MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

(2)

ENTATION PAGE

Form Approved
OMB No. 0704-0188

AD-A264 600

...ORT DATE	3. REPORT TYPE AND DATES COVERED
	Reprint IEEE August 1992

4. TITLE AND SUBTITLE
Multiscale Autoregressive Processes, Part II: Lattice Structures for Whitening and Modeling

5. FUNDING NUMBERS
DAAL03-86-K-0171

6. AUTHOR(S)
M. Basseville
A. Benveniste
A.S. Willsky

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dr. Sanjoy K. Mitter 24635-MA-UIR
M.I.T.
Laboratory for Information and Decision Systems
Cambridge, MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 24635.432-MA-UIR

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

DTIC
S MAY 0 3 1993 E

13. ABSTRACT (Maximum 200 words)

Abstract—In part I of this two-part paper we introduced a class of stochastic processes defined on dyadic homogenous trees. The motivation for the study of these processes comes from our desire to develop a theory for multiresolution descriptions of stochastic processes in one and multiple dimensions based on the idea underlying the recently introduced theory of wavelet transforms. In part I we described how this objective leads to the study of processes on trees and began the development of a theory of autoregressive (AR) models for isotropic processes on trees. In this second part we complete that investigation by developing lattice structures for the whitening and modeling of isotropic processes on trees. We also present a result relating the stability properties of these models to the reflection coefficient sequence introduced in part I. In addition, this framework allows us to obtain a detailed analysis of the Wold decomposition of processes on trees. One interesting aspect of this is that there is a significantly larger class of singular processes on dyadic trees than on the integers.

93-09807

410 950

SUBJECT TERMS

93 5 05 111

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRAC
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Multiscale Autoregressive Processes, Part II: Lattice Structures for Whitening and Modeling

Michèle Basseville, Albert Benveniste, *Fellow, IEEE*, and Alan S. Willsky, *Fellow, IEEE*

***Abstract*—In part I of this two-part paper we introduced a class of stochastic processes defined on dyadic homogenous trees. The motivation for the study of these processes comes from our desire to develop a theory for multiresolution descriptions of stochastic processes in one and multiple dimensions based on the idea underlying the recently introduced theory of wavelet transforms. In part I we described how this objective leads to the study of processes on trees and began the development of a theory of autoregressive (AR) models for isotropic processes on trees. In this second part we complete that investigation by developing lattice structures for the whitening and modeling of isotropic processes on trees. We also present a result relating the stability properties of these models to the reflection coefficient sequence introduced in part I. In addition, this framework allows us to obtain a detailed analysis of the Wold decomposition of processes on trees. One interesting aspect of this is that there is a significantly larger class of singular processes on dyadic trees than on the integers.**

I. Introduction

In part I [1] of this two-part paper we introduced the class of isotropic processes on homogeneous dyadic trees, and began the analysis of the corresponding class of autoregressive (AR) processes. As developed in [1], the motivation for the study of these processes comes from our desire to provide a statistical framework for multiscale signal processing based on the structure of the recently introduced class of wavelet transforms [7].

In [1], we introduced and described the geometry of homogeneous dyadic trees and a natural notion of "past" and "future," where a move into the "past" ("future") corresponds to moving to a coarser (finer) scale description of a signal. The class of isotropic processes on trees was also introduced in [1], and, with our notions of past and future, we defined the class of autoregressive (AR) isotropic processes and began the study of their parametrization. The major result of [1] was to establish that the only suitable parametrization of isotropic processes on the dyadic tree is obtained via reflection coefficients following the generalization of the Schur–Levinson parametrization techniques for usual time series. In this second part, we further investigate the properties of isotropic processes in terms of their reflection coefficients. In particular, in this paper we use the analysis in [1] both to construct lattice structures for the whitening and modeling of AR processes on dyadic trees and to analyze in detail these models and the properties of isotropic processes.

This paper relies heavily on the framework and results of [1], and we refer the reader to that paper for reference. In the next section we provide a brief summary of some of the basic notation and constructs from [1]. Section III is then devoted to the presentation of whitening and modeling filters for AR isotropic processes. Unnormalized as well as normalized versions of these filters are given. In particular, the normalized modeling filter appears as a tree structured scattering system. Then, in Section IV, several properties of isotropic processes are analyzed in terms of the reflection coefficient sequence. Specifically, AR processes are characterized as being the processes with only finitely many nonzero reflection coefficients, purely nondeterministic processes are characterized in a fairly simple way, a stability result for the modeling filters is presented, and finally it is shown that every finite set of reflection coefficients properly define a unique AR process provided they belong to an easily defined domain. Finally, future issues, both practical and theoretical, are discussed in the conclusion. Many of the results presented here, while paralleling those for time series, are more complex than their time series counterparts due to the significant increase in geometric complexity in going from a homogeneous tree of order 1, i.e., the usual discrete-time index set, to the dyadic tree, which is of order 2. For example, as introduced in [1] and described in detail in Section IV, the prediction error processes associated with lattice filters on dyadic trees are vector processes of dimension that increases with filter order.

II. Dyadic Trees, Isotropic Processes, and Prediction Error Recursions

In this section we review some of the basic concepts and constructs described in [1]. We refer the reader to [1] for details.

Manuscript received March 20, 1990; revised May 30, 1991. The work of M. Basseville and A. Benveniste was supported in part by Grant CNRS G0134 and by INRIA-NSF agreement. The work of A. S. Willsky was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-92-J-0002 and by the National Science Foundation under Grants MIP-9015281 and INT-9002393, and in part by the U.S. Army Research Office under Contract DAAL03-86-K-0171, and by INRIA.

M. Basseville is with IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France, and with the Centre National de la Recherche Scientifique (CNRS), France.

A. Benveniste is with IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France, and with the Institut National de Recherche en Informatique et en Automatique (INRIA), France.

A. S. Willsky is with the Laboratory for Information and Decision Systems, and the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

IEEE Log Number 9201075.

1053-587X/92$03.00

A homogeneous dyadic tree $\mathfrak{I}$, as illustrated in Fig. 1, has a natural notion of distance $d(s, t)$ between any two nodes, $s, t \in \mathfrak{I}$. By choosing a particular boundary point, denoted by $-\infty$, we can redraw $\mathfrak{I}$ as in Fig. 2. Here all of the points that are at the same "distance from $-\infty$" appear on the same level or horocycle. For multiscale processing we can think of each horocycle as corresponding to describing signals at a particular scale, with finer scales being farther from $-\infty$. Also, as illustrated in the figure the choice of $-\infty$ leads naturally to a backward (fine-to-coarse) shift, γ^{-1}, and two forward (coarse-to-fine) shifts α and β. Also it is useful to introduce the operator δ. As indicated in Fig. 2, the transformation δ which interchanges nodes t and $t\delta$ for all $t \in \mathfrak{I}$, can be thought of locally as an interchange pivoted at the immediate ancestor $t\gamma^{-1}$. Higher order operators $\delta^{(n)}$ correspond to interchanges pivoted at more distant ancestors of t (i.e., $t\gamma^{-n}$, $n > 1$). The nodes $t\delta^{(2)}$ and $t\delta^{(3)}$ are indicated in the figure.

As developed in [1], all nodes in $\mathfrak{I}$ can be coded in terms of shifts from a specified, arbitrary node t_0. Specifically, let

$$\mathfrak{L} = (\gamma^{-1})^* \cup (\gamma^{-1})^* \delta\{\alpha, \beta\}^* \cup \{\alpha, \beta\}^*. \quad (2.1)$$

Then $\mathfrak{I} = \{t_0 w | w \in \mathfrak{L}\}$. The order $|w|$ of any move $w \in \mathfrak{L}$ is defined as

$$|w| = d(t, tw). \quad (2.2)$$

A move w is causal, denoted by $w \leq 0$, if wt is on the same or a coarser horocycle than that on which t is located.

A zero-mean stochastic process Y_t, $t \in \mathfrak{I}$, indexed by nodes on the tree is isotropic if the correlation between Y at any two nodes depends only on the distance between those nodes, i.e.,

$$E\{Y_t Y_s\} = r_{d(t,s)}. \quad (2.3)$$

Equivalently, Y_t is isotropic if $Z_t = Y_{f(t)}$ has the same statistics as Y_t for any isometry f: $\mathfrak{I} \to \mathfrak{I}$, i.e., any one to one and onto map of $\mathfrak{I}$ onto itself that preserves distances. An AR model of order p has the form

$$Y_t = \sum_{\substack{w \leq 0 \\ |w| \leq p}} a_w Y_{tw} + \sigma W_t \quad (2.4)$$

where W_t is unit variance white noise. Our interest here is in developing AR models for isotropic processes, and as discussed in [1], the constraints of isotropy imply rather complex constraints on the a_w coefficients in (2.4). Note also that the number of these coefficients essentially doubles as the order increases by 1.

In [1] we began the process of developing an alternate description of isotropic AR processes in terms of generalizations of the Levinson and Schur recursions for stationary time series. Because of the structure of the dyadic trees, in particular the fact that the number of nodes at a given distance from a specified node increases geometrically with distance, the development of these recursions

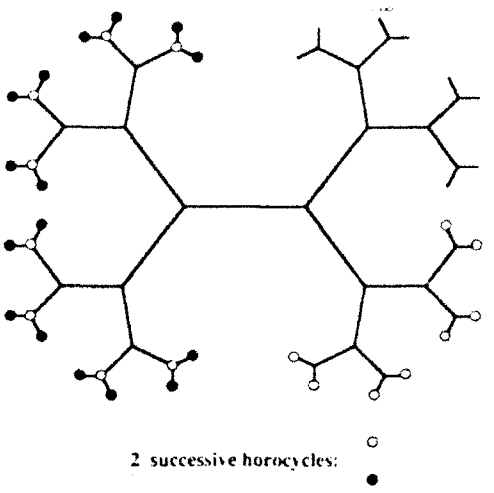

Fig. 1. A homogenous dyadic tree, with a choice for the boundary point ∞ indicated

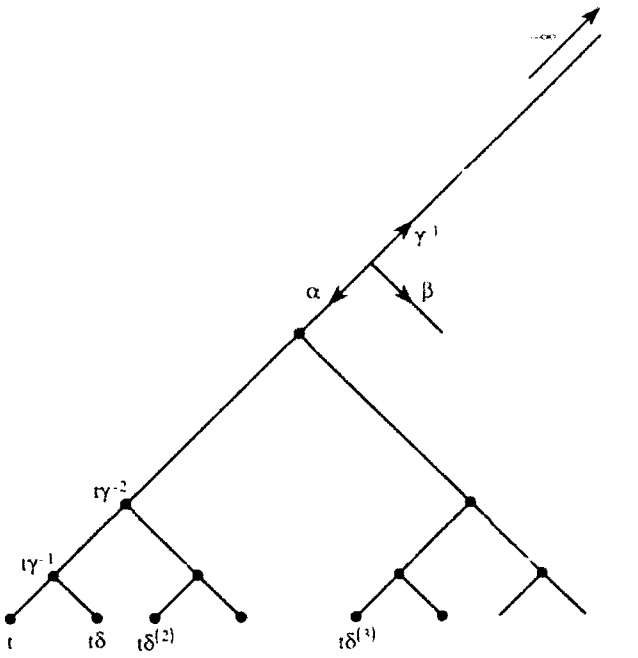

Fig. 2. Redrawing the dyadic tree with a particular choice of boundary point $-\infty$

and the lattice filters to be described here involve prediction error vectors of dimensions increasing with the order of prediction. Specifically, define the nth order past of Y at node t

$$\mathcal{Y}_{t,n} = \mathcal{H}\{Y_{tw}: w \leq 0, |w| \leq n\} \quad (2.5)$$

where $\mathcal{H}\{\cdots\}$ denotes the linear span of a set of random variables. Then the nth order backward prediction errors at node t are given by

$$\mathcal{F}_{t,n} = \mathcal{H}\{F_{t,n}(w): |w| = n, w \leq 0\} \quad (2.6)$$

where

$$F_{t,n}(w) = Y_{tw} - E(Y_{tw} | \mathcal{Y}_{t,n-1}). \quad (2.7)$$

We also let $F_{t,n}$ denote the full 2^{n-2}-dimensional vector

of $F_{t,n}(w)$, where $[x]$ = largest integer $\leq x$ and where the ordering of the w in (2.6) used in constructing $F_{t,n}$ is described in [1]. Similarly, we have the nth order forward prediction errors at node t

$$\mathcal{E}_{t,n} \triangleq \mathcal{K}\{E_{t,n}(w): |w| < n \text{ and } w \neq 0\} \tag{2.8}$$

where

$$E_{t,n}(w) \triangleq Y_{tw} - E(Y_{tw} \mid \mathcal{Y}_{t_{\gamma^{n-1}},n-1}). \tag{2.9}$$

The ordering of the w in (2.8) to construct the $2^{[(n-1)/2]}$-dimensional vector $E_{t,n}$ is described in [1].

In [1] we began the analysis of the recursive computation of these prediction errors as the order n increases. What we found was that as in the usual Levinson recursions for time series the forward and backward prediction errors of one order could be expressed recursively in terms of projections onto prediction error vectors of the preceding order. Most importantly, the constraint of isotropy allowed us to show that the required projections onto multidimensional spaces such as $\mathcal{F}_{t,n}$ and $\mathcal{E}_{t,n}$ reduced in fact to projections onto specific scalar random variables, namely, the barycenters of the prediction error vectors:

$$e_{t,n} = 2^{-[(n-1)/2]} \sum_{|w|<n, w\neq 0} E_{t,n}(w) \tag{2.10}$$

$$f_{t,n} = 2^{-[n/2]} \sum_{|w|=n, w\leq 0} F_{t,n}(w). \tag{2.11}$$

Indeed, these projections can be expressed recursively in terms of a single scalar sequence of reflection coefficients k_n. Furthermore, as shown in [1], there exist a set of scalar Levinson recursions for the barycenter error processes. In particular, for n even

$$e_{t,n} = e_{t,n-1} - k_n f_{t\gamma^{n-1},n-1} \tag{2.12}$$

$$f_{t,n} = \tfrac{1}{2}\left(f_{t\gamma^{n-1},n-1} + e_{t\delta^{(n-2)},n-1}\right) - k_n e_{t,n-1} \tag{2.13}$$

where

$$k_n = \operatorname{cor}\left(e_{t,n-1}, f_{t\gamma^{n-1},n-1}\right) \tag{2.14}$$

and $\operatorname{cor}(x, y) = E(xy)/[E(x^2)E(y^2)]^{1/2}$. For n odd, $n > 1$:

$$e_{t,n} = \tfrac{1}{2}\left(e_{t,n-1} + e_{t\delta^{((n-1)/2)},n-1}\right) - k_n f_{t\gamma^{n-1},n-1} \tag{2.15}$$

$$f_{t,n} = f_{t\gamma^{n-1},n-1} - \tfrac{1}{2} k_n \left(e_{t,n-1} + e_{t\delta^{((n-1)/2)},n-1}\right) \tag{2.16}$$

with

$$k_n = \operatorname{cor}\left(\tfrac{1}{2}\left(e_{t,n-1} + e_{t\delta^{((n-1)/2)},n-1}\right), f_{t\gamma^{n-1},n-1}\right). \tag{2.17}$$

Also, for $n = 1$, $e_{t,1} = E_{t,1}$, $f_{t,1} = F_{t,1}$, and

$$F_{t,1} = Y_{t\gamma^{0}} - k_1 Y_t \tag{2.18}$$

$$E_{t,1} = Y_t - k_1 Y_{t\gamma^{0}} \tag{2.19}$$

where

$$k_1 = \frac{E[Y_{t\gamma^{0}} Y_t]}{E[Y_{t\gamma^{0}}^2]} = \frac{r_1}{r_0}. \tag{2.20}$$

In addition, the variances of the prediction errors satisfy the following: for n even

$$\sigma_{e,n}^2 = E(e_{t,n}^2) = (1 - k_n^2)\sigma_{n-1}^2 \tag{2.21}$$

$$\sigma_{f,n}^2 = E(f_{t,n}^2) = \left(\frac{1 + k_n}{2} - k_n^2\right)\sigma_{n-1}^2. \tag{2.22}$$

For n odd

$$\sigma_{e,n}^2 = \sigma_{f,n}^2 = \sigma_n^2 = (1 - k_n^2)\sigma_{n-1}^2 \tag{2.23}$$

where (2.23) holds for $n = 1$ as well, with $\sigma_0^2 = r_0$. By using these equations it is possible to derive a recursive procedure for computing the k_n that is the counterpart of the recursions in the standard Levinson algorithm for time series. We also have Schur recursions which provide an alternative mechanism for computing the reflection coefficient sequence. Specifically, define the formal power series

$$P_n = \operatorname{cov}(Y_t, e_{t,n}) = \sum_{w\leq 0} E(Y_t e_{tw,n}) \cdot w \tag{2.24}$$

$$Q_n = \operatorname{cov}(Y_t, f_{t,n}) = \sum_{w\leq 0} E(Y_t f_{tw,n}) \cdot w \tag{2.25}$$

and recall the following operators on formal power series we introduced in [1]: given

$$S = \sum_{w\in\mathcal{L}} s_w \cdot w$$

we set

$$\gamma[S] = \sum_{w\in\mathcal{L}} s_{w\gamma^{0}} \cdot w$$

$$\delta^{(k)}[S] = \sum_{w\in\mathcal{L}} s_{w\delta^{(k)}} \cdot w.$$

Then for n even

$$P_n = P_{n-1} - k_n \gamma[Q_{n-1}] \tag{2.26}$$

$$Q_n = \tfrac{1}{2}\left(\gamma[Q_{n-1}] + \delta^{(n-2)}[P_{n-1}]\right) - k_n P_{n-1} \tag{2.27}$$

where

$$k_n = \frac{\gamma[Q_{n-1}](0) + \delta^{(n-2)}[P_{n-1}](0)}{2P_{n-1}(0)} \tag{2.28}$$

while for n odd

$$P_n = \tfrac{1}{2}\left(P_{n-1} + \delta^{((n-1)/2)}[P_{n-1}]\right) - k_n \gamma[Q_{n-1}] \tag{2.29}$$

$$Q_n = \gamma[Q_{n-1}] - k_n \tfrac{1}{2}\left(P_{n-1} + \delta^{((n-1)/2)}[P_{n-1}]\right) \tag{2.30}$$

$$k_n = \frac{2\gamma[Q_{n-1}](0)}{P_{n-1}(0) + \delta^{((n-1)/2)}[P_{n-1}](0)} \tag{2.31}$$

where

$$P_0 = Q_0 = \sum_{w\leq 0} r_w \cdot w. \tag{2.32}$$

We also note here that, as for time series, there are constraints on the reflection coefficients, which, thanks to the conditions required for isotropy, are slightly more complex for isotropic processes on dyadic trees:

$$\text{for } n \text{ even}, \quad -\tfrac{1}{2} \leq k_n \leq 1 \tag{2.33}$$

$$\text{for } n \text{ odd}, \qquad -1 \le k_n \le 1. \tag{2.34}$$

As we develop in this paper, these results lead to lattice structures for AR processes on dyadic trees in which only one new reflection coefficient is introduced as the order increases by one. Furthermore, the constraints (2.33), (2.34) on these coefficients are quite simple and are decoupled from one another. Thus the lattice filter parametrization of AR processes is far superior to the direct AR model (2.4).

III. Vector Levinson Recursions and Modeling and Whitening Filters

In [1] we showed that the recursive computation of the components of the prediction error vectors $E_{t,n}$ and $F_{t,n}$ involved projections onto the barycenter error processes. In addition, we developed scalar Levinson recursions for the barycenters. In this section we combine these results in order to develop whitening and modeling filters for Y_t. As we will see, in order to produce true whitening filters, it will be necessary to perform a further normalization of the innovations. However, the formulas for $E_{t,n}$ and $F_{t,n}$ are simpler, and consequently we begin with them.

A. Filters Involving the Unnormalized Residuals

To begin, let us introduce a variation on notation used to describe the structure of the covariance matrix of the prediction error $E_{t,n}$ which we denoted in [1] by $\Sigma_{E,n}$. In particular, we let 1_* denote a unit vector all of whose components are the same:

$$1_* = \frac{1}{\sqrt{\dim 1}}\, 1. \tag{3.1}$$

We also define the matrix

$$U_* = 1_* 1_*^T \tag{3.2}$$

which has a single nonzero eigenvalue of 1. Equations (3.1), (3.2) define a family of vectors and matrices of different dimensions. The dimension used in any of the expressions to follow is that required for the expression to make sense. We also note the following identities:

$$U_* U_* = U_* \tag{3.3}$$

$$f^* = 1_*^T F = \frac{1}{\sqrt{\dim F}} \sum_w F(w) \tag{3.4}$$

$$1f = 1_* f^* = U_* F \tag{3.5}$$

where $F = \{F(w)\}$ is a vector indexed by words w ordered as described in [1], where f is its barycenter, and where f^* is a normalized version of its barycenter.

The results of [1, Sec. IV] lead directly to the following recursions for the prediction error vectors:

Theorem 3.1: The prediction error vectors $E_{t,n}$ and $F_{t,n}$ satisfy the following recursions, where the k_n are the reflection coefficients for the process Y_t:

For n even:

$$E_{t,n} = E_{t,n-1} - k_n U_* F_{t,n-1} \tag{3.6}$$

$$F_{t,n} = \begin{vmatrix} E_{t0,n-1} \\ F_{t1,n-1} \end{vmatrix} - k_n \begin{vmatrix} U_* \\ U_* \end{vmatrix} E_{t,n-1} \tag{3.7}$$

For n odd, $n > 1$:

$$E_{t,n} = \begin{vmatrix} E_{t0,n-1} \\ E_{t1,n-1} \end{vmatrix} - k_n U_* F_{t,n-1} \tag{3.8}$$

$$F_{t,n} = F_{t1,n-1} - k_n U_* \begin{vmatrix} E_{t0,n-1} \\ E_{t1,n-1} \end{vmatrix} \tag{3.9}$$

while for $n = 1$ $F_{t,1}$ and $E_{t,1}$ are scalars satisfying (2.18), (2.19). Here the reflection coefficient sequence k_n is calculated from the correlation function, r_n, of Y_t according to either the Levinson or Schur recursions described in Section II.

Proof: Equations (2.18), (2.19) for $n = 1$ are exactly [1, eqs. (3.17), (3.19)]. As indicated previously, the remainder of this result is also a direct consequence of the analysis in [1, sec. 3 and 4]. For example, from (3.16), [1, lemma 4.1, eq. (4.6)], and (3.5) of this paper, we have the following chain of equalities for n even:

$$\begin{aligned} E_{t,n} &= E_{t,n-1} - E(E_{t,n-1} | F_{t,n-1}) \\ &= E_{t,n-1} - \lambda 1 f_{t,n-1} \\ &= E_{t,n-1} - \lambda U_* F_{t,n-1} \end{aligned} \tag{3.10}$$

where λ is a constant to be determined. If we premultiply this equality by $(\dim E_{t,n-1})^{-1} 1^T$, we obtain the formula for the barycenter of $E_{t,n-1}$, and from (2.12) we see that $\lambda = k_n$. The other formulas are obtained in an analogous fashion.

The form of these whitening filters deserves some comment. Note first that the stages of the filter are of growing dimension, reflecting the growing dimension of the $E_{t,n}$ and $F_{t,n}$ as n increases. Nevertheless, each stage is characterized by a single reflection coefficient. Thus, while the dimension of the innovations vector of order n is on the order of 2^{n-2}, only n coefficients are needed to specify the whitening filter for its generation. This, of course, is a direct consequence of the constraint of isotropy and the richness of the group of isometries of the tree.

In [1] we obtained recursions (2.21)–(2.23) for the variances of the barycenters of the prediction vectors. Theorem 3.1 above provides us with the recursions for the covariances and correlations for the entire prediction error vectors. We summarize these and other facts about these covariances in the following.

Corollary. Let $\Sigma_{E,n}$, $\Sigma_{F,n}$ denote the covariances of $E_{t,n}$ and $F_{t,n}$, respectively. Then

1) For n even:

a) The eigenvalue of $\Sigma_{E,n}$ associated with the eigen-

vector $[1, \cdots, 1]$ is

$$\mu_{E,n} = 2^{(n-2)-1}\sigma_{e,n}^2 \tag{3.11}$$

where $\sigma_{e,n}^2$ is the variance of $e_{t,n}$.

b) The eigenvalue of $\Sigma_{E,n}$ associated with the eigenvector $[1, \cdots, 1]$ is

$$\mu_{E,n} = 2^{n-2}\sigma_{f,n}^2 \tag{3.12}$$

where $\sigma_{f,n}^2$ is the variance of $f_{t,n}$.

2) For n odd:

$$\Sigma_{E,n} = \Sigma_{F,n} = \Sigma_n \tag{3.13}$$

and the eigenvalue associated with the eigenvector $[1, \cdots, 1]$ is

$$\mu_n = \mu_{E,n} = \mu_{F,n} = 2^{(n-1)-2}\sigma_n^2 \tag{3.14}$$

where σ_n^2 is the variance of both $e_{t,n}$ and $f_{t,n}$.

3) For n even:

$$\Sigma_n = \Sigma_{F,n} = \operatorname{cov}\begin{pmatrix} E_{t,n} \\ E_{t\delta^{(n-2)},n} \end{pmatrix} = \begin{vmatrix} \Sigma_{E,n} & \lambda_n U \\ \lambda_n U & \Sigma_{E,n} \end{vmatrix} \tag{3.15}$$

where $U = 11^T$, and

$$\Sigma_{E,n} = \Sigma_{n-1} - k_n^2\sigma_{n-1}^2 U \tag{3.16}$$

$$\lambda_n = (k_n - k_n^2)\sigma_{n-1}^2. \tag{3.17}$$

4) For n odd, $n > 1$:

$$\Sigma_n = \begin{vmatrix} \Sigma_{E,n-1} & \lambda_{n-1} U \\ \lambda_{n-1} U & \Sigma_{E,n-1} \end{vmatrix} - k_n^2\sigma_{f,n-1}^2 U. \tag{3.18}$$

5) For $n = 1$:

$$\Sigma_1 = (1 - k_1^2) r_0. \tag{3.19}$$

Proof: Equations (3.11), (3.12), and (3.14) follow directly from the definition of the barycenter. For example, for n even

$$2^{(n/2)-1} e_{t,n} = 1^T E_{t,n} \tag{3.20}$$

from which (3.11) follows immediately. Equations (3.13) is a consequence of [1, lemma 4.1]. To verify (3.15) let us first evaluate (3.6) at both t and $t\delta^{(n-2)}$:

$$\begin{pmatrix} E_{t,n} \\ E_{t\delta^{(n-2)},n} \end{pmatrix} = \begin{pmatrix} E_{t,n-1} \\ E_{t\delta^{(n-2)},n-1} \end{pmatrix} - k_n \begin{pmatrix} U_* \\ U_* \end{pmatrix} F_{t\gamma_{n-1},n-1}. \tag{3.21}$$

The first equality in (3.15) is then a direct consequence of [1, lemma 4.1] (compare (3.7) and (3.21)). The form given in the rightmost expression in (3.15) is also immediate: the equality of the diagonal blocks is due to isotropy, while the form of the off-diagonal blocks again follows from [1, lemma 4.1]. The specific expression for $\Sigma_{E,n}$ in (3.16) follows directly from the second equality in (3.10), while (3.17) follows from (3.21) and the fact that

$$E[E_{t,n-1}(w) E_{t\delta^{(n-2)},n-1}(w')] = k_n\sigma_{n-1}^2 \tag{3.22}$$

which in turn follows from [1, lemma 4.1 and (4.27)]. Finally, (3.18) follows from (3.15) and (3.8), and (3.19) is immediate from (2.18)–(2.20).

Just as with time series, the whitening filter specification leads directly to a modeling filter for Y.

Theorem 3.2: The modeling filter for Y_t is given by the following. For n even

$$\begin{pmatrix} E_{t,n-1} \\ F_{t,n} \end{pmatrix} = S(k_n) \begin{pmatrix} E_{t,n} \\ E_{t\delta^{(n-2)},n} \\ F_{t\gamma_{n-1},n-1} \end{pmatrix} \tag{3.23}$$

where[1]

$$S(k_n) = \begin{bmatrix} I & 0 & k_n U_* \\ -k_n U_* & I & (k_n - k_n^2) U_* \\ -k_n U_* & 0 & (I - k_n^2 U_*) \end{bmatrix}. \tag{3.24}$$

For n odd, $n > 1$:

$$\begin{pmatrix} E_{t,n-1} \\ E_{t\delta^{(n-2)},n-1} \\ F_{t,n} \end{pmatrix} = S(k_n) \begin{pmatrix} E_{t,n} \\ F_{t\gamma_{n-1},n-1} \end{pmatrix} \tag{3.25}$$

where

$$S(k_n) = \begin{bmatrix} I & k_n U_* \\ -k_n U_* & (I - k_n^2 U_*) \end{bmatrix} \tag{3.26}$$

while for $n = 1$:

$$\begin{pmatrix} Y_t \\ F_{t,1} \end{pmatrix} = \begin{pmatrix} 1 & k_1 \\ -k_1 & 1 - k_1^2 \end{pmatrix} \begin{pmatrix} E_{t,1} \\ Y_{t\gamma_1} \end{pmatrix} = S(k_1) \begin{pmatrix} E_{t,1} \\ Y_{t\gamma_1} \end{pmatrix}. \tag{3.27}$$

These equations can be verified by solving (3.6)–(3.9) and (2.18)–(2.20) to obtain expressions for E's of order $n-1$ and F's of order n in terms of E's of order n and F's of order $n-1$. Note should also be made of the dimensions of the various signals and matrices in Theorem 3.2. In particular, for n even the two components on the left-hand side of (3.23) are of dimensions $2^{(n-2)-1}$ and $2^{(n-2)}$, respectively, while all three of the vectors on the right-hand side of (3.23) are of dimension $2^{(n-2)-1}$ and each of the square blocks in (3.24) is $2^{(n-2)-1}$-dimensional. For n odd, $n > 1$, both components of the right-hand side of (3.25) are $2^{(n-2)-2}$-dimensional as is the $F_{t,n}$-block on the left-hand side. The two E-blocks on the left-hand side, however, are $2^{(n-3)-2}$-dimensional, and the

[1] In fact, we should properly write $S(k_n, n)$ since the dimension of the blocks depends on n; nevertheless, we choose to write $S(k_n)$ to simplify the notation; this will be done everywhere in the sequel.

four square blocks in (3.26) are $2^{(N-1)/2}$-dimensional. We have included dotted lines in (3.23)–(3.26) to emphasize how these mappings operate. Note also that the first section (3.27) of the modeling filter involves only scalar quantities.

As is the case for time series, the lattice modeling filter of Theorem 3.2 has a scattering layer structure. An important difference here is that the growing dimension of the prediction errors leads to a tree-like structure for the scattering diagram, and because of this, we find that groups of values of Y are calculated together in this structure. In particular, from Theorem 3.2 we can deduce that if we consider a modeling filter of odd length N, then this modeling filter can be viewed as a map from the $2^{(N-1)/2}$-dimensional input vector $E_{t,N}$ to the $2^{(N-1)/2}$-dimensional set of outputs $\{Y_{t\delta^w} \| w| \le N, w \ne 0\}$. For N even, the modeling filter maps the two $2^{(N-2)/2}$-dimensional input vectors $E_{t,N}$, $E_{t\delta^{N/2},N}$ to the $2^{(N-2)/2}$-dimensional set of outputs $\{Y_{t\delta^w} \| w| \le N, w \ne 0\}$. The case of $N = 6$ is illustrated in Fig. 3. In this case the input vectors $E_{t,6}$ and $E_{t\delta^3,6}$ produce the outputs $Y_{t\delta^w}$ for $w \ne 0$, $|w| \le 6$ (as well as the backward errors $F_{t,6}$ and $F_{t\delta^3,6}$, which are not actually needed for the recursion). The E vectors of various orders propagate from left to right, while the F's propagate from right to left. The small black squares represent γ^{-1} operations and the blocks labeled "1," "2," etc., perform the computations described in Theorem 3.2. The details of the operation of this system, however, requires further explanation.

Let us first look at the situation for n odd, in which case each block labeled "n" performs the calculations given by (3.25) (or (3.27) for $n = 1$). For example, the inputs to the top "3" block in the figure (which has been shaded) are $E_{t,3}$ and $F_{t\delta^{-1},2}$, while the outputs are $E_{t,2}$, $E_{t\delta,2}$, and $F_{t,3}$. Note that this block is connected to the right to systems generating both Y_t and $Y_{t\delta}$, but apparently we do not need a corresponding "3" block at $t\delta$ in addition to the one at t. To understand this, consider writing (3.25) at $t\delta^{(m-1)\cdot 2^n}$ rather than at t:

$$\begin{bmatrix} E_{t\delta^{(m-1)\cdot 2^n},n-1} \\ E_{t,n-1} \\ \\ F_{t\delta^{(m-1)\cdot 2^n},n} \end{bmatrix} = S(k_n) \begin{bmatrix} E_{t\delta^{(m-1)\cdot 2^n},n} \\ F_{t\delta^{(m-1)\cdot 2^n+\gamma_n,n-1}} \end{bmatrix} \quad (3.28)$$

where we have used the fact that for any k, $\delta^{(k)}$ is its own inverse. Note that the first two components of the output in (3.28) are simply a permutation of the first two in (3.25). The last outputs in these equations and both inputs apparently differ. However, it is easily checked that the outputs $F_{t,n}$ and $F_{t\delta^{(m-1)\cdot 2^n},n}$ are identical up to a permutation of the ordering of components, as are the input pair $E_{t,n}$ and $E_{t\delta^{(m-1)\cdot 2^n},n}$ and the input pair $F_{t\gamma_n,n-1}$ and $F_{t\delta^{(m-1)\cdot 2^n+\gamma_n,n-1}}$ (this latter fact is proved in the "umbilical lemma" of Appendix A and expressed via the "umbilical cords"—dotted connections—of Fig. 3). Thus there is actually no need to have a "3" block at $t\delta$ as there was at t, or more generally an "n" block at $t\delta^{(m-1)\cdot 2^n+\gamma_n}$ as well as at t.

For n even, the blocks labeled "n" perform the calculations as specified by (3.23). For example, the inputs to the top "2" box in the figure are $F_{t,2}$, $E_{t,2}$, and $F_{t\gamma_2,1}$, while the outputs are $E_{t,3}$ and $F_{t,2}$. Again it is important to examine the analogous computation at a related point. Specifically, consider evaluating (3.23) at the point $t\delta^{(m-2)\cdot 2^n}$:

$$\begin{bmatrix} E_{t\delta^{(m-2)\cdot 2^n},n+1} \\ \\ F_{t\delta^{(m-2)\cdot 2^n},n} \end{bmatrix} = S(k_n) \begin{bmatrix} E_{t\delta^{(m-2)\cdot 2^n},n} \\ E_{t\delta^{(m-2)\cdot 2^n+\gamma_n},n-1} \\ F_{t\delta^{(m-2)\cdot 2^n+\gamma_n},n} \end{bmatrix} \quad (3.29)$$

Note first that the first outputs of (3.23) and (3.29), namely, $E_{t,n+1}$ and $E_{t\delta^{(m-2)\cdot 2^n},n+1}$ are in fact distinct, and thus it is necessary to implement the computation (3.29). For example, the second "2" block (also shaded) in Fig. 3 computes as one of its outputs $E_{t\delta^2,3}$. Next note that the other outputs, $F_{t,n}$ and $F_{t\delta^{(m-2)\cdot 2^n},n}$ of (3.23) and (3.29) are not identical. However, these signals must pass through a γ_n^{-1} operation before entering the corresponding "$n+1$" block, and we have already seen in our analysis of the odd case that $F_{t,n}$ and $F_{t\delta^{(m-2)\cdot 2^n},n}$ are identical up to a permutation. Thus only one of these is needed for a block at level $n+1$. This is indicated in the figure by a connecting, dotted bar between the γ_n^{-1} block immediately to the left of pairs of even numbered blocks, with only one of these identical signals continuing backward to the corresponding $n+1$ block. For example, the two left-going output signals of the shaded "2" blocks, $F_{t,2}$ and $F_{t\delta^2,2}$, are merged in this way after the γ_2^{-1} operation on each.

Examining next the right-hand sides of (3.23) and (3.29) we see that the first two inputs are identical except for a flip in the order. This is captured in the figure, as can be seen for $n = 2$, where the two inputs entering from the left of the second "2" block are the same as those for the first "2" block, except in reverse position. It is also not difficult to check that the last inputs $F_{t,n-1}$ and $F_{t\delta^{(m-2)\cdot 2^n+\gamma_n,n-1}}$ are identical up to a permutation of components. While these signals do enter individual blocks we have again indicated that they are the same by a connecting dotted bar between the γ_n^{-1} blocks immediately to the right of pairs of even numbered blocks. For the case of $n = 2$, the two left-going input signals of the shaded "2" blocks, $F_{t,1}$ and $F_{t\delta,1}$ are identical and are connected by such a dotted bar.

Figs. 4 and 5 describe in more explicit terms the data flow and memory structure for the system of order 6. Specifically suppose that we have finished the computations required at the horocycle indicated with squares in Fig. 4. As indicated in the tree at the top of this figure (via a shaded bar connecting the squares), sets of 4 nodes at this level are coupled together (more generally for an nth-order model $2^{(n-1)-(n-2)}$ points are coupled together). The state for this set of four nodes is indicated above the nodes: we have stored the scalar values of Y, F_1, and E_3 at each node, we have stored the 2-vectors E_2 and F_2 for each of two

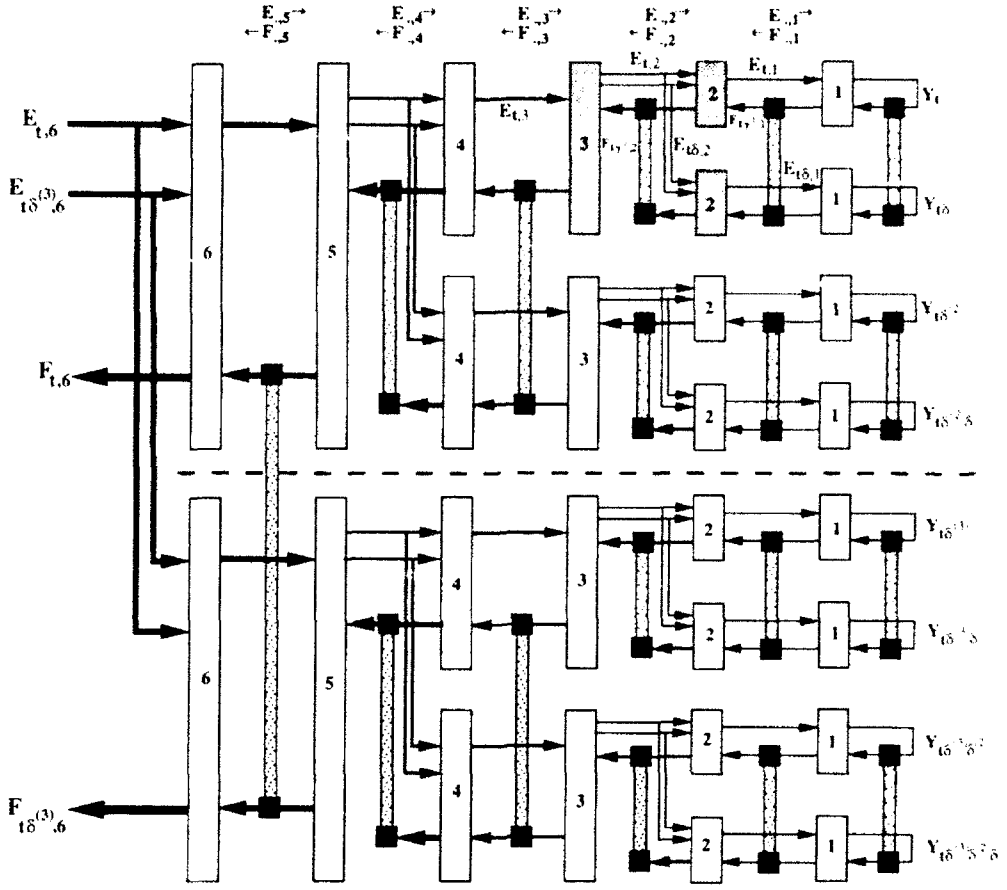

Fig. 3. Illustrating the scattering lattice structure of the modeling filter of Theorem 3.2 for a sixth-order model. Each block labeled "n" performs the computation in (3.23) (for n even), (3.25) (for n odd, $n > 1$), or (3.27) (for $n = 1$). The small solid squares denote γ^{-1} operations, and the dotted connections between such squares (the umbilical cords) indicate signals (at the outputs of these squares), that are identical up to a permutation of components. As indicated at the top of the figure, the signals flowing through this system are the E and F error processes of successive orders, with the E's flowing left to right and the F's right to left.

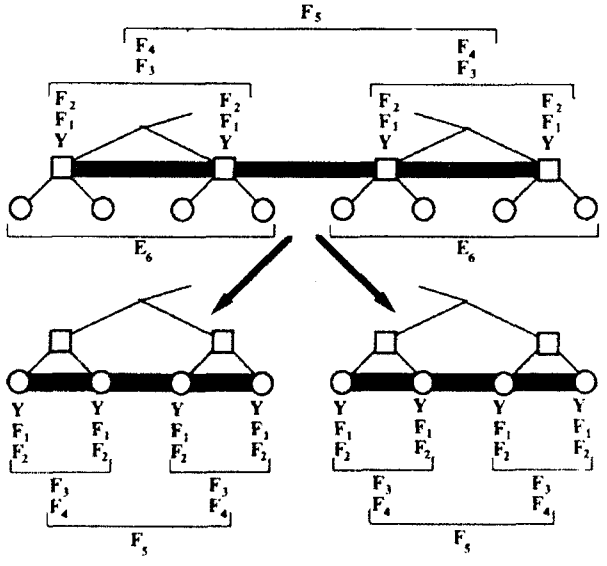

Fig. 4. Illustrating the propagation of state information for the filter in Fig. 3. The stored information (indicated above the top portion of the figure) for a set of four nodes at the horocycle indicated by squares is used, together with the input E_6 vectors, to compute the two corresponding sets of information at the two descendent groups of four points at the next horocycle.

pairs of these nodes, and we have stored a single 4-vector F_5 for the set of 4 nodes. Given these quantities and the two 4-vector E_6 inputs for each of the two sets of 4 descendent nodes (indicated with circles, with a connecting bar for each set), the model performs two parallel computations (which are identical in structure) to produce the required variables to be stored at each of the two sets of descendant nodes. Fig. 5 illustrates in more detail how these computations are distributed and performed. Here at each level the variables required as inputs are indicated with "?," while those produced as outputs are indicated with "!." Furthermore those inputs corresponding to the stored state are indicated above each layer of the computation, while below each figure we indicate the inputs received externally ($?E_6$) or from previous layers (all other $?E$'s). We also indicate below each layer the outputs produced, some of which (the $!F$'s at layers 2–6 and the $!Y$'s at level 1) form components of the state at the next horocycle while others of which (the $!E$'s) are used as inputs by succeeding layers. For example, at the top level F_5 is stored and two E_6 vectors are received as the only external inputs. This layer, as shown in Fig. 3, has two actual sets of outputs. One of these, the F_6 vectors, is not needed for

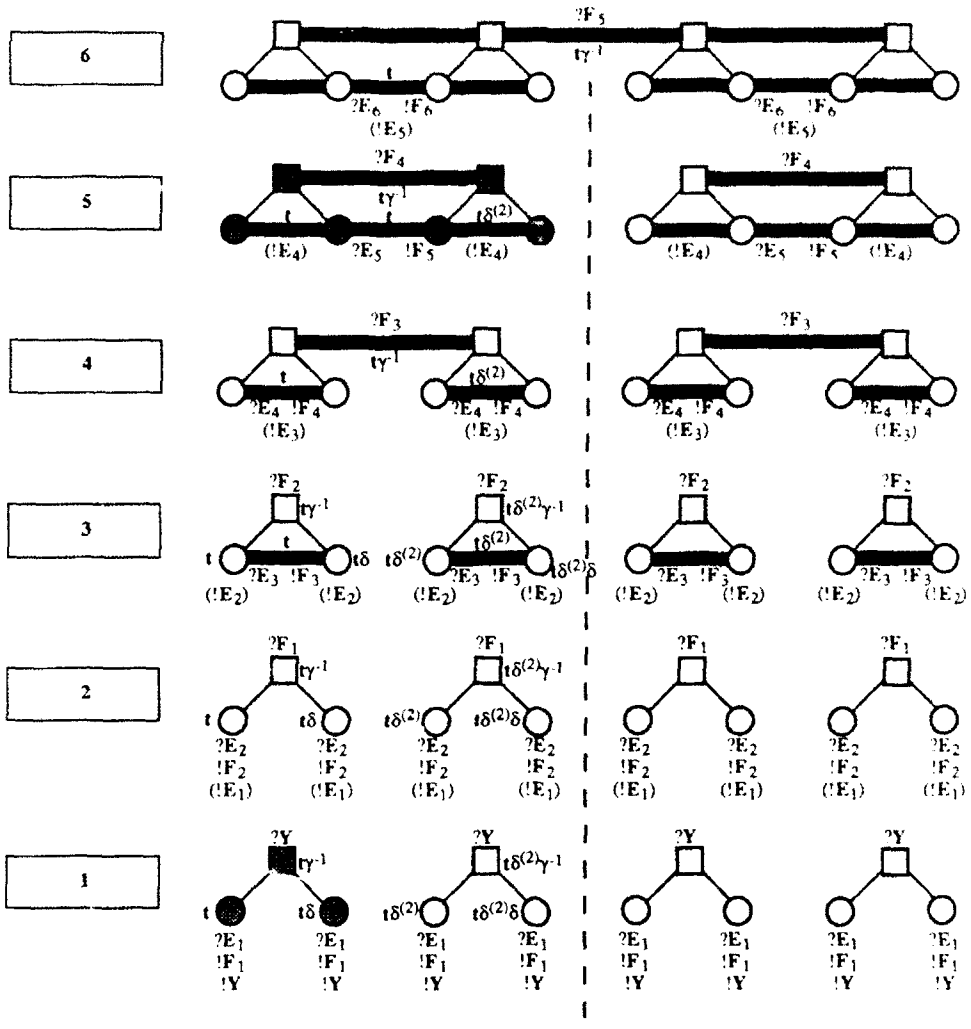

Fig. 5. Illustrating the detailed computational flow for the propagation of state information for the filter described in Figs. 3 and 4.

the subsequent computation and indeed is typically not computed in lattice implementations. The other outputs produced are the E_5 vectors which will not be stored as part of the state at the next horocycle but which do show up as inputs to the layer 5 blocks.

We have also included node indices in part of Fig. 5 to make it easier to connect the computational structure of the figure with the computations described in (3.23)–(3.27). For example, the lower left-hand portion of layer 1 (distinguished by shaded circles and [illegible]tures) corresponds to the pair of computations corresponding to (3.27) evaluated at t and at $t\delta$. Also, at higher layers, we encounter vector error processes, and as we have seen, these vectors are not distinct or, in fact, needed at all nodes. For example, consider the portion of the layer 5 computations indicated by shaded squares. This describes the computation of (3.25) for $n = 5$, which requires a single E_5 input at node t, a single stored F_4 vector at $t\gamma^{-1}$ and produces one F_5 vector at node t and two E_4 vectors at t and $t\delta^{(2)}$. In this case, as we have pointed out, a single F_4 vector needs to be stored for the pair of square nodes connected by the solid bar in the figure. We have indicated its index $t\gamma^{-1}$ in the center of the bar. Similarly, the index t of the single E_5 and F_5 vectors is indicated in the center of the lower solid bar, while the indices, t and $t\delta^{(2)}$, for the two E_4 vectors are indicated above the appropriate portion of the solid bar. Note that the apparent redundancies, indicated by the shaded bars in Fig. 4, are not present in Fig. 5, as in this figure we have shown just those variables required to be stored and computed from horocycle to horocycle.

As we will see, understanding the structure of the filter described in Figs. 3–5 greatly facilitates our analysis of stability.

B. Levinson Recursions for the Normalized Residuals

The prediction errors $E_{t,n}$ and $F_{t,n}$ do not quite define isotropic processes. In particular, the components of these vectors representing prediction error vectors at a set of nodes are correlated. Furthermore, for n even we have seen that $E_{t,n}$ and $E_{t\delta^{(2)},n-1}$ are correlated (see (3.15)). These observations provide the motivation for the normalized recursions developed in this section. In this de-

velopment we use the superscript * to denote normalized versions of random vectors. Specifically, $X^* = \Sigma_X^{-1/2} X$ where Σ_X is the covariance of X and $\Sigma_X^{1/2}$ is its symmetric, positive definite square root.

We now can state and prove the following.

Theorem 3.3: For n odd the covariance matrix Σ_n defined in (3.13) is invertible if and only if $-1 < k_n < 1$. For n even, Σ_n as defined in (3.15), is invertible if and only if $-1/2 < k_n < 1$. Under these conditions the whitening recursions of Theorem 3.1 can be normalized, yielding the following recursions for the normalized residuals:

For n even:

$$\begin{pmatrix} E_{t,n} \\ E_{t\delta^{(n-2)},n} \end{pmatrix}^* = \Theta(k_n) \left| \begin{pmatrix} E^*_{t,n-1} \\ E^*_{t\delta^{(n-2)},n-1} \end{pmatrix} - k_n \begin{pmatrix} U_* \\ U_* \end{pmatrix} F^*_{t\gamma^{-1},n-1} \right| \tag{3.30}$$

$$F^*_{t,n} = \Theta(k_n) \left| \begin{pmatrix} E^*_{t\delta^{(n-2)},n-1} \\ F^*_{t\gamma^{-1},n-1} \end{pmatrix} - k_n \begin{pmatrix} U_* \\ U_* \end{pmatrix} E^*_{t,n-1} \right| \tag{3.31}$$

where $\Theta^{-1}(k_n)$ is the matrix square root satisfying[2]

$$\Theta^{-2}(k_n) = \begin{pmatrix} I - k_n^2 U_* & (k_n - k_n^2) U_* \\ (k_n - k_n^2)_* & I - k_n^2 U_* \end{pmatrix}. \tag{3.32}$$

For n odd, $n > 1$

$$E^*_{t,n} = \Theta(k_n) \left| \begin{pmatrix} E_{t,n-1} \\ E_{t\delta^{(n-1)/2},n-1} \end{pmatrix}^* - k_n U_* F^*_{t\gamma^{-1},n-1} \right| \tag{3.33}$$

$$F^*_{t,n} = \Theta(k_n) \left| F^*_{t\gamma^{-1},n-1} - k_n U_* \begin{pmatrix} E_{t,n-1} \\ E_{t\delta^{(n-1)/2},n-1} \end{pmatrix}^* \right| \tag{3.34}$$

where

$$\Theta^{-2}(k_n) = I - k_n^2 U_*. \tag{3.35}$$

For $n = 1$:

$$E^*_{t,1} = \frac{1}{\sqrt{1 - k_1^2}} (Y^*_t - k_1 Y^*_{t\gamma^{-1}}) \tag{3.36}$$

$$F^*_{t1} = \frac{1}{\sqrt{1 - k_1^2}} (Y^*_{t\gamma^{-1}} - k_1 Y^*_t). \tag{3.37}$$

Remark: Note that for n even we normalize $E_{t,n}$ and $E_{t\delta^{(n-2)},n}$ together as one vector, while for n odd, $E_{t,n}$ is normalized individually. This is consistent with the nature of their statistics as described in (3.15)–(3.19) and with the fact that for n even $\dim F_{\cdot,n} = 2 \dim E_{\cdot,n}$, while for n odd $\dim F_{\cdot,n} = \dim E_{\cdot,n}$.

Proof: Let us first derive (3.30)–(3.37) assuming the invertibility of Σ_n for each n. This result is a relatively straightforward computation given (3.11)–(3.19). For n even we begin with (3.7) and (3.21) and premultiply each by

$$\operatorname{diag}(\Sigma_{n-1}^{-1/2}, \Sigma_{n-1}^{-1/2}).$$

Since 1_* is an eigenvector of Σ_{n-1}, Σ_{n-1} and therefore $\Sigma_{n-1}^{-1/2}$ commute with U_*. This immediately yields (3.30) and (3.31) where the matrix $\Theta(k_n)$ is simply the inverse of the square root of the covariance of the term in brackets in (3.30) and in (3.31) (the equality of these covariances follows from (3.15)). Equation (3.32) then follows from (3.11) and (3.15). The case of n odd involves an analogous set of steps, and the $n = 1$ case is immediate.

The preceding analysis provides us both with the conditions for the invertibility of Σ_n and with a recursive procedure for calculating $\Sigma_n^{-1/2}$ (see [1, appendix D] for an alternate efficient procedure). For n even we have

$$\Sigma_n^{-1/2} = \Theta(k_n) \operatorname{diag}(\Sigma_{n-1}^{-1/2}, \Sigma_{n-1}^{-1/2}) \tag{3.38}$$

while for n odd, $n > 1$

$$\Sigma_n^{-1/2} = \Theta(k_n) \Sigma_{n-1}^{-1/2} \tag{3.39}$$

and for $n = 1$

$$\Sigma_1^{-1/2} = [(1 - k_1^2) r_0]^{-1/2}. \tag{3.40}$$

Note first that from (3.40) we see that $|k_1|$ must be less than 1 for $\Sigma_1^{-1/2}$ to exist. For $n > 1$ and odd, note that the only nonunity eigenvalue of $I - k_n^2 U_*$ is $1 - k_n^2$, and thus $\Theta(k_n)$ exists for n odd if and only if $k_n^2 < 1$. Also in this case we can readily compute $\Theta(k_n)$ using the following formula. For any $k > -1$

$$(I + kU_*)^{-1/2} = I + \left(\frac{1}{\sqrt{1 + k}} - 1 \right) U_* \tag{3.41}$$

For n even, we make use of the result that for S and T symmetric

$$\begin{pmatrix} S & T \\ T & S \end{pmatrix}^{-1/2} = \frac{1}{2} \begin{pmatrix} X + Y & X - Y \\ X - Y & X + Y \end{pmatrix} \tag{3.42}$$

where

$$X = (S + T)^{-1/2}$$
$$Y = (S - T)^{-1/2}. \tag{3.43}$$

Using (3.42), (3.43) we see from (3.32) that to calculate $\Theta(k_n)$ for n even we must calculate

$$(I + (k_n - 2k_n^2) U_*)^{-1/2}$$

and

$$(I - k_n U_*)^{-1/2}$$

which exist if and only if $-1/2 < k_n < 1$, completing our proof.

[2] Again, to be precise, we should write $\Theta(k_n, n)$ rather than $\Theta(k_n)$. For simplicity we use the less cumbersome notation.

If $k_n = -1/2$ or 1 for n even or $k_n = +1$ for n odd, the resulting error processes are not full rank. This is the simplest example of a singular process, for which perfect prediction of a linear combination of Y's on a given horocycle can be obtained using only a finite set of values of Y on "past" horocycles. In Section IV we will characterize the full class of singular processes in terms of its infinite reflection coefficient sequence.

Now that we have a normalized form for the residual vectors, we can also describe the normalized version of the modeling filters which provide the basis for generating isotropic Y_t's specified by a finite number of reflection coefficients and driven by white noise.

Theorem 3.4: A normalized modeling filter for the isotropic process Y_t exists if and only if $-1 < k_n < 1$ for n odd and $-1/2 < k_n < 1$ for n even. In this case, this filter has the following form. For n even we have

$$\begin{pmatrix} E^*_{t,n-1} \\ F^*_{t,n} \end{pmatrix} = \Sigma(k_n) \begin{pmatrix} \begin{pmatrix} E_{t,n} \\ E_{t\delta^{n-1},n} \end{pmatrix}^* \\ F^*_{t\delta,n-1} \end{pmatrix} \tag{3.44}$$

where[3]

$$\Sigma(k_n) = \begin{pmatrix} I + a(k_n)U_* & b(k_n)U_* & k_nU_* \\ -\frac{k_n}{2}U_* & I + c(k_n)U_* & b(k_n)U_* \\ d(k_n)U_* & -\frac{k_n}{2}U_* & I + a(k_n)U_* \end{pmatrix} \tag{3.45}$$

with

$$a(k) = \frac{\sqrt{1+2k}+1}{2}\sqrt{1-k} - 1 \tag{3.46}$$

$$b(k) = \frac{\sqrt{1+2k}-1}{2}\sqrt{1-k} \tag{3.47}$$

$$c(k) = \frac{\sqrt{1+2k}-(1+k)}{2} \tag{3.48}$$

$$\hat{\Sigma}(k_n) = \begin{bmatrix} O^{-1}(k_n) & k_n \begin{pmatrix} U_* \\ U_* \end{pmatrix} \\ O(k_n)\begin{pmatrix} -k_nU_* & I \\ -k_nU_* & 0 \end{pmatrix} O^{-1}(k_n) & O(k_n)\begin{pmatrix} (k_n - k_n^2)U_* \\ I - k_n^2U_* \end{pmatrix} \end{bmatrix}. \tag{3.58}$$

$$d(k) = -c(k) - k. \tag{3.49}$$

The matrix $\Sigma(k_n)$ is referred to as the scattering matrix, and it satisfies

$$\Sigma(k_n)\Sigma^J(k_n) = I. \tag{3.50}$$

[3]Again we shorten the notation and write $\Sigma(k_n)$ rather than $\Sigma(k_n, n)$.

For n odd, $n \neq 1$

$$\begin{pmatrix} E^*_{t,n-1} \\ F^*_{t,n} \end{pmatrix} = \Sigma(k_n) \begin{pmatrix} E^*_{t,n} \\ F^*_{t\delta,n-1} \end{pmatrix} \tag{3.51}$$

where the scattering matrix

$$\Sigma(k_n) = \begin{pmatrix} (I - k_n^2U_*)^{1/2} & k_nU_* \\ -k_nU_* & (I - k_n^2U_*)^{1/2} \end{pmatrix} \tag{3.52}$$

satisfies

$$\Sigma(k_n)\Sigma^J(k_n) = I. \tag{3.53}$$

For $n = 1$:

$$\begin{pmatrix} Y^*_t \\ F^*_{t,1} \end{pmatrix} = \Sigma(k_1)\begin{pmatrix} E^*_{t,1} \\ Y^*_{t\delta} \end{pmatrix} \tag{3.54}$$

and

$$\Sigma(k_1) = \begin{pmatrix} \sqrt{1-k_1^2} & k_1 \\ -k_1 & \sqrt{1-k_1^2} \end{pmatrix} \tag{3.55}$$

also satisfies

$$\Sigma(k_1)\Sigma^J(k_1) = I. \tag{3.56}$$

Proof: We begin by solving (3.30) for $(E^{*J}_{t,n-1}, E^{*J}_{t\delta^{n-1},n-1})^J$ then by substituting this into (3.31) we obtain

$$\begin{bmatrix} \begin{pmatrix} E^*_{t,n-1} \\ E^*_{t\delta^{n-1},n-1} \end{pmatrix}^* \\ F^*_{t,n} \end{bmatrix} = \hat{\Sigma}(k_n) \begin{bmatrix} \begin{pmatrix} E_{t,n} \\ E_{t\delta^{n-1},n} \end{pmatrix}^* \\ F^*_{t\delta,n-1} \end{bmatrix} \tag{3.57}$$

where

To obtain the desired relation, we simply drop the calculation of $E^*_{t\delta^{n-1},n-1}$ from (3.57). To do this explicitly we consider $\hat{\Sigma}(k_n)$ as a matrix with three block columns and four block rows (one each for $E^*_{t,n-1}$ and $E^*_{t\delta^{n-1},n-1}$ and two for $F^*_{t,n}$). Thus what we wish to do is to drop the second block row. A careful calculation using the relations derived previously yields (3.45)–(3.49). That $\Sigma(k_n)$ satisfies (3.50) follows immediately from the fact that the

vectors on both sides of (3.44) have identity covariances. The result for n odd, $n > 1$ is obtained in a similar fashion, and the case of $n = 1$ is immediate.

IV. Reflection Coefficients and the Properties of Processes and Models

The analysis in [1] and in the preceding sections provides us with a framework in which we can say a great deal about stochastic processes and dynamic systems on trees. In the first subsection we provide a complete characterization of isotropic autoregressive processes, and in Subsection IV-B we characterize purely nondeterministic processes. In Section IV-C we relate the stability of the lattice models on trees to the reflection coefficients, while in Section IV-D we show that all lattice filters with appropriately-constrained reflection coefficients yield AR processes, showing the one-to-one correspondence between these filters and processes. In each case there are similarities to the analysis for stationary time series. However, the more complex structure of the dyadic tree leads to some important and substantive differences.

A. Characterization of Autoregressive Processes

A well-known and essentially trivial result for time series is that if Y_t is a pth order autoregressive process, then the reflection coefficients k_n are 0 for $n \geq p + 1$. Furthermore, the pth-order forward and backward prediction errors, which are also identical to the nth order prediction errors for $n \geq p + 1$, form white noise sequences. The following result, which states the counterpart of this result for isotropic processes on trees, requires some prefatory comment. Specifically, thanks to the vector nature of our models, i.e., the fact that a group of Y's on a given horocycle are generated together from a group of the E's, the prediction error processes whose whiteness we consider consist of sampled versions of the (normalized) E and F processes, with one "sample" taken per "group." In particular, from our discussion at the end of Section III-A and from the definition (2.8), (2.9) of $E_{t,n}$, we find that the components of $E_{t,n}$ and $E_{tw,n}$ are permutations of one another if $w \neq 0$ and $|w| \leq n - 1$. Thus we need only consider one of these vectors for each group on each horocycle. Note that this means that we are choosing only one out of $2^{(n-1)2^n}$ error vectors, but each vector is exactly of dimension $2^{(n-1)2^n}$ so that we do have the correct number of total degrees of freedom, one per node on the tree.

Turning to the backward residuals, we find from the discussion in Section III-A and the definition (2.6), (2.7) of $F_{t,n}$ that the components of $F_{t,n}$ and $F_{tw,n}$ for $w \neq 0$ and $|w| \leq n - 1$ are permutations of one another. On the other hand, as pointed out (for $w = \delta^{(n-2)}$ in Section III-A, if n is even, so that it is possible to find $w \neq 0$ with $|w| = n$, $F_{t,n}$ and $F_{tw,n}$ do not have identical sets of components. Furthermore, it is easily checked that these vectors are not uncorrelated. However, as is also pointed out in Section III-A, the signals $F_{t_1,n}$ and $F_{tw_1,n}$ do have identical component sets, and it is only these "delayed" signals that play a role in the modeling filter. Thus for our purposes here we need choose only one vector from the set $\{F_{tw,n}; w \neq 0, |w| \leq n\}$. In this case we are choosing one $2^{(n-2)2^n}$-dimensional vector from a set of $2^{(n-2)2^n}$ such vectors, again producing the correct number of degrees of freedom.

Finally, as we have noted in Section III-B, it is necessary to normalize the prediction error processes. For the backward prediction errors, this simply means that we will consider the $F^*_{t,n}$ rather than the $F_{t,n}$. Similarly for n odd we consider the $E^*_{t,n}$. However for n even our normalization involves the combined normalization of $E_{t,n}$ and $E_{t\delta^{n-1},n}$ (e.g., referring to Fig. 3, the two inputs $E_{t,n}$ and $E_{t\delta^{n-1},n}$ are normalized together). Thus for n even, instead of choosing one vector from $\{E_{tw,n}; w \neq 0, |w| \leq n - 1\}$ $\{E_{t\delta^{n-1}w,n}; w \neq 0, |w| \leq n - 1\}$ we choose one vector (of twice the dimension) from

$$\left\{ \begin{pmatrix} E_{tw,n} \\ E_{t\delta^{n-1}w,n} \end{pmatrix}^* \middle| w \neq 0, |w| \leq n - 1 \right\}.$$

Proposition 4.1: If Y_t is an AR(p) isotropic process, then the reflection coefficients k_n are 0 for $n \geq p + 1$. Furthermore, the forward and backward normalized prediction error vectors of order p and greater form standard white noise processes (i.e., with unity covariance). More precisely, let t_0 be an arbitrary node on the tree, and consider an infinite sequence of predecessors and successors to t_0:

$$T = \{t_0\gamma^{-k} | k \geq 0\} \cup \{t_0\alpha^k | k \geq 1\}.$$

Then for any $n \geq p$, the set of backward prediction error vectors

$$\left\{ F^*_{t\delta^{n-1},n} | t \in T, j > \frac{n}{2} \right\}$$

forms a standard white noise process. Similarly, for $n \geq p$ and odd, the set

$$\left\{ E^*_{ts,n} | s \in T, j > \frac{n}{2} \right\}$$

forms a standard white noise process, while for $n \geq p$ and even, the set

$$\left\{ \begin{pmatrix} E_{ts,n} \\ E_{ts\delta^{n-1},n} \end{pmatrix}^* \middle| s \in T, j > \frac{n}{2} \right\}$$

forms a standard white noise process.

The construction of T and the choices of points forming the sets of prediction errors in proposition 4.1 represents one particular way of choosing one prediction error vector from each of the sets described before the statement of the proposition.

Proof of Proposition 4.1: We focus explicitly on the E's, as an analogous proof holds for the F's. Note first that, thanks to the normalization, all of the E^* variables do have unity covariance. Also, thanks to the sampling

done in forming the E^* sets, it is straightforward to check that the whiteness will be proven if we can show that for $n \geq p$ (and either even or odd) the unnormalized prediction error $E_{t,n}$ is uncorrelated with $E_{tw,n}$ (denoted $E_{t,n} \perp E_{tw,n}$) for $w < 0$ and for $w \neq 0$, $|w| > n$.

Showing that this is true for $w < 0$ is essentially the same as the proof in the time series case. Specifically, if $k_n = 0$ for $n \geq p$, then,

$$E_{t,2m} = E_{t,2m-1} \qquad \text{if } 2m > p$$

$$E_{t,2m+1} = \begin{vmatrix} E_{t,2m} \\ E_{tw^{-1},2m} \end{vmatrix} \qquad \text{if } 2m \geq p \tag{4.1}$$

so that, $n \geq p$,

$$E_{t,n} \perp \mathcal{Y}_{t\gamma^{-1},\infty} \tag{4.2}$$

by definition of the forward prediction errors. Hence from (2.9) we see that for $n \geq p$,

$$E_{t,n} \perp E_{tw,n} \qquad \text{for } w < 0. \tag{4.3}$$

Hence it remains to prove that

$$E_{t,n} \perp E_{t\delta^{-j},n} \qquad \text{for } j > \frac{n}{2}. \tag{4.4}$$

This proof, which involves the construction of isometries much as in several of the proofs in [1], is sketched in Appendix C.

For a time series model the constraint of causality severely restricts the support of its impulse response. For example any AR time series model has an AR impulse response whose support is the nonnegative integers. For processes on trees, however, there is considerable flexibility in the possible choice of support for a causal impulse response. However, as the following states, the constraints of isotropy allow us to determine precisely the support for AR models.

Proposition 4.2: Let Y_t be an AR (p) isotropic process. Let us write the formal power series P_p defined in (2.24) as

$$P_p = \sum_{\substack{w \in \mathbb{L} \\ w \leq 0}} p_w \cdot w. \tag{4.5}$$

If $p = 0$, $p_w = 0$ if $w \neq 0$. If $p = 1$, $p_w = 0$ unless $w = \gamma^{-k}$ for some $k \geq 0$. If $p \geq 2$, then $p_w = 0$ for all words of the form $w = \gamma^{-k}\delta w_{\alpha\beta}$ with

$$w_{\alpha\beta} \in \{\alpha, \beta\}^* \qquad \text{and } |w_{\alpha\beta}| > \left|\frac{p}{2}\right| - 1. \tag{4.6}$$

In other words, P_p has its support in a cylinder of radius $\lfloor p/2 \rfloor$ around the path $\{\gamma^{-k}\}$ toward $-\infty$. From this we also have that the modeling filter of an AR (p) process has its support in the same cylinder of radius $\lfloor p/2 \rfloor$ around $[t, -\infty) = \{t\gamma^{-k} | k \geq 0\}$. Conversely, any process such that the modeling filter has its support contained in the cylinder of radius $\lfloor p/2 \rfloor$ is necessarily an AR (p) process.

Comment: The proof of this result is straightforward, although tedious, and is left to the reader. Fig. 6 illustrates the cylinders for low order AR processes. Note that proposition 4.2 is a generalization of the result in [1, appendix A] which states that if an isotropic process has its support concentrated on $[t, -\infty)$, then it is necessarily AR (1).

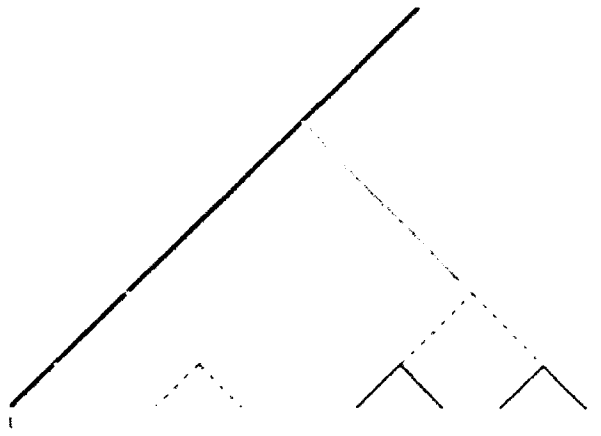

Fig. 6. Illustrating the cylinder of radius 0 (support of AR (1): dark solid lines), cylinder of radius 1 (support of AR (2): AR (3): dark solid and gray shaded lines), and cylinder of radius 2 (support of AR (4): AR (5): dark solid, gray shaded, and dashed lines).

B. Characterization of Regular (or Purely Nondeterministic) Processes

Definition 4.1: We shall say that an isotropic process Y_t is regular or purely nondeterministic if

$$\sigma^2 > 0 \tag{4.7}$$

holds, where

$$\sigma^2 = \inf \left\| \left(\sum_{w \leq 0} \mu_w Y_{tw} \right) - E\left(\left(\sum_{w \leq 0} \mu_w Y_{tw} \right) \middle| \mathcal{Y}_{t\gamma^{-1},\infty} \right) \right\|^2 \tag{4.8}$$

and the infimum ranges over all collections of scalars $(\mu_w)_{w \leq 0}$ where only finitely many of the μ_w are nonzero and the condition $\Sigma \mu_w^2 = 1$ is satisfied.

In other words, no nonzero linear combination of the values of Y_t on any given horocycle can be predicted exactly with the aid of knowledge of Y in the strict past, $\mathcal{Y}_{t\gamma^{-1},\infty}$ and the associated prediction error is uniformly bounded from below. We shall now characterize regular processes in a fairly simple way using reflection coefficients.

Theorem 4.1: i) The following formulas hold for every isotropic process:

$$\sigma^2 = \lim_{n \to \infty} \inf \lambda_{\min}(\Sigma_{2n+1}) \tag{4.9}$$

$$\lambda_{\min}(\Sigma_{2n+1}) = r_0(1 - k_1^2) \prod_{p=1}^{n} (1 - k_{2p-1}^2) \cdot \min\{1 + k_{2p} - 2k_{2p}^2, 1 - k_{2p}\} \tag{4.10}$$

where $\lambda_{\inf}(A)$ denotes the smallest eigenvalue of the matrix A, and Σ_{2n+1} is defined in (3.13).

ii) An isotropic process Y_t is regular if and only if its reflection coefficient sequence is such that $|k_{2n+1}| < 1$, $-\frac{1}{2} < k_{2n} < +1$, and furthermore,

$$\sum_{n=1}^{\infty} (k_{2n+1}^2 + |k_{2n}|) < \infty. \tag{4.11}$$

Comment: The corresponding characterization of regular processes in the case of time series is (cf., for instance, [2]):

$$|k_n| < 1 \ \forall n, \qquad \sum_{n=1}^{\infty} k_n^2 < \infty. \tag{4.12}$$

Proof: Note first that the singularity of the process if $|k_{2n+1}| = 1$ or if $k_n = -1/2$ or 1 follows directly from the resulting degeneracy of the prediction error covariance (Theorem 3.3). Condition (4.11) of point ii is an immediate consequence of point i, since for k small enough $(1 - k, 1 + k - 2k^2) \sim 1 - |k|$. Thus we shall only prove i. First, let us prove (4.9) by showing that σ^2 is both $\geq$ and $\leq$ the right-hand side of (4.9). With every $(\mu_w)_{w \neq 0}$ as in definition 4.1 we associate a sequence of vectors (M_n) of increasing dimension. Specifically, we begin by forming an infinite-dimensional vector by ordering the μ_w according to the ordering on the $w \neq 0$ defined in [1, sec. III-A]. For each n we then take the vector $\tilde{M}_n$ to be the truncated version of this infinite vector by keeping only the initial segment consisting of those μ_w's such that w is involved in the definition of $E_{t,2n+1}$. We then set $M_n = \tilde{M}_n / \|\tilde{M}_n\|$ if $\tilde{M}_n \neq 0$, and equal to some arbitrary unit vector otherwise (here, $\| \cdots \|$ denotes the usual Euclidian norm).

We obviously have

$$\tilde{M}_n = M_n, \qquad \text{for } n \text{ large enough.} \tag{4.13}$$

Hence, thanks to the limit theorem for square integrable martingales [8], [9], we can write, for the considered family (μ_w)

$$\begin{aligned}
&\left\| \sum_{w \neq 0} \mu_w Y_{tw} - E\left(\sum_{w \neq 0} \mu_w Y_{tw} \,|\, \mathcal{Y}_{t\gamma_{n-1},\infty} \right) \right\|^2 \\
&\quad = \lim_{n \to \infty} \left\| \sum_{w \neq 0} \mu_w Y_{tw} - E\left(\sum_{w \neq 0} \mu_w Y_{tw} \,|\, \mathcal{Y}_{t\gamma_{n-1},2n} \right) \right\|^2 \\
&\quad = \lim_{n \to \infty} M_n^T \Sigma_{2n+1} M_n \\
&\quad \geq \liminf_{n \to \infty} \lambda_{\inf}(\Sigma_{2n+1})
\end{aligned} \tag{4.14}$$

where the second equality uses (4.13), and the inequality is due to the fact that M_n is a unit vector. Since the last expression in (4.14) does not involve the considered family (μ_w), we immediately get the inequality $\geq$ in (4.9). Now, fix $\epsilon > 0$ and select n_ϵ large enough so that $\lambda_{\inf}(\Sigma_{2n_\epsilon+1}) - \epsilon$ is smaller than the right-hand side of (4.9). Then, take for M_{n_ϵ} a unit eigenvector of $\Sigma_{2n_\epsilon+1}$ associated with its smallest eigenvalue. We then obtain following inequalities which, since ϵ is arbitrary, yields the inequality $\leq$ in (4.9):

$$\begin{aligned}
&\liminf_{n \to \infty} \lambda_{\inf}(\Sigma_{2n_\epsilon+1}) + \epsilon \\
&\quad \geq \lambda_{\inf}(\Sigma_{2n_\epsilon+1}) \\
&\quad = M_{n_\epsilon}^T \Sigma_{2n_\epsilon+1} M_{n_\epsilon} \\
&\quad \geq \left\| \sum_{w \neq 0} \mu_w^{n_\epsilon} Y_{tw} - E\left(\sum_{w \neq 0} \mu_w^{n_\epsilon} Y_{tw} \,|\, \mathcal{Y}_{t\gamma_{n_\epsilon-1},2n_\epsilon} \right) \right\|^2 \\
&\quad \geq \sigma^2
\end{aligned}$$

where $(\mu_w^{n_\epsilon})$ is the family associated with M_{n_ϵ}.

It remains to prove (4.10). Using (3.38)–(3.40), we can write

$$\Sigma_{2n+1}^{(1,2)} = \Theta(k_{2n+1})\Theta(k_{2n}) \begin{pmatrix} \Sigma_{2n-1}^{(1,2)} & 0 \\ 0 & \Sigma_{2n-1}^{(2,1)} \end{pmatrix}$$

But the three matrices on the right-hand side of this formula all have the Haar system as eigenvectors (cf. [1, eq. (4.19)]. Hence we can diagonalize all of these matrices simultaneously:

$$\begin{aligned}
\Lambda(\Sigma_{2n+1}^{(1,2)}) = \Lambda(\Theta_{2n+1}^*)\Lambda(\Theta_{2n}^*) \\
\cdot \begin{pmatrix} \Lambda(\Sigma_{2n-1}^{(1,2)}) & 0 \\ 0 & \Lambda(\Sigma_{2n-1}^{(2,1)}) \end{pmatrix}
\end{aligned}$$

holds, where $\Lambda(A)$ denotes the diagonal matrix of the eigenvalues of A. Using the definition of $\Theta^{-1}(k,\ell)$ in (3.32), (3.35), we can deduce that

$$\begin{aligned}
\Lambda(\Sigma_{2n+1}) = \operatorname{diag}(1 - k_{2n+1}^2, 1 \cdots 1) \\
\cdot \operatorname{diag}(1 + k_{2n} - 2k_{2n}^2, 1 - k_{2n}, 1 \cdots 1) \\
\cdot \begin{pmatrix} \Lambda(\Sigma_{2n-1}) & 0 \\ 0 & \Lambda(\Sigma_{2n-1}) \end{pmatrix}
\end{aligned}$$

so that, by expanding the product and using (3.36), we finally get (4.10). This finishes the proof of the theorem.

Note that the condition (4.11) is much more easily violated by a valid reflection coefficient sequence than the corresponding expression (4.12) for time series, pointing to the fact that there is apparently a far richer class of singular processes on trees than on the real line. This is apparently related to the characterization of spectral measures for isotropic processes and to the large size of the boundary of the dyadic tree (see the comments concerning [1, eq. (2.33)]).

C. A Stability Criterion

A well-known result for all-pole lattice filters is that such a filter is stable if and only if all of the reflection coefficients have magnitude less than 1. In this section we

state and prove Theorem 4.2, which is the counterpart of this result for the lattice filters introduced in this paper. Before stating this result, let us clarify what we mean by "stability." Figs. 3–5 depict (for a sixth-order example) the structure of the unnormalized filter. This filter describes how the computation of Y_i propagates from horocycle to horocycle, with $E_{i,n}$ (for n odd) or $(E_{i,n}, E_{i\delta^{n/2},n})$ (for n even) as input and the corresponding block of Y's on the same horocycle as output. It is the stability of this filter that we wish to study.

Theorem 4.2: Under the conditions

$$-1 < k_n < 1 \qquad n \text{ odd } 1 \le n \le N \tag{4.15}$$

$$-\tfrac{1}{2} < k_n < 1 \qquad n \text{ even } 1 \le n \le N \tag{4.16}$$

the Nth-order unnormalized modeling filter specified by (3.23)–(3.27) is stable, so that a bounded input $E_{i,N}$ (for N odd) or $(E_{i,N}, E_{i\delta^{N/2},N})$ (for N even) yields a bounded output Y_i. Similarly, the normalized modeling filter specified in Theorem 3.4 is also stable under these conditions so that a bounded input $E_{i,N}^*$ (for N odd) or $(E_{i,N}, E_{i\delta^{N/2},N})^*$ (for N even) yields a bounded output Y_i^*.

Proof: Let us first show that by taking advantage of the structure of the filter computations we can simplify the required analysis and can, in fact, reduce it to a question of stability analysis for a standard temporal system. To begin, in Fig. 7 we have depicted one of the two parallel computations depicted in Fig. 5, where we have used notation that emphasizes the sequential nature of the computations. Here the indices "m" and "$m-1$" index horocycles so that the "$m-1$" quantities are stored and the "m" quantities are computed from the input $(E_{61}(m), E_{62}(m))$ which is distinguished by a solid box at level 6 in the figure (note that the reverse-going output from this final level, $F_{61}(m)$ is distinguished by a dashed box). The subscripts for the signals in Fig. 7 code the various error and output processes at each level. The first subscript for the E and F vectors indicate the order of the error vector, while the second subscript (and the only subscript for the Y's) indexes the vectors along a segment of a horocycle. The precise correspondence between the normalized version of quantities in Fig. 7 and those in Fig. 5 can be directly determined by matching up signals and node indices in Fig. 5 with signals and horocycles index (m and $m-1$) in Fig. 7. For example,

$$Y_i,\ Y_{i\delta},\ Y_{i\delta^2},\ Y_{i\delta^2\delta} \leftrightarrow Y_1(m),\ Y_2(m),\ Y_3(m),\ Y_4(m)$$

$$Y_{i\gamma_1},\ Y_{i\delta^2\gamma_1} \leftrightarrow Y_1(m-1),\ Y_2(m-1)$$

$$F_{i,3},\ F_{i\delta^2,3} \leftrightarrow F_{31}(m),\ F_{32}(m)$$

$$E_{i,4},\ E_{i\delta^2,4} \leftrightarrow (E_{41}(m),\ E_{42}(m)).$$

As we emphasized in Section III-A, and as illustrated graphically in Figs. 3–5 and 7, each stage of the computation is pyramidal in structure. For example, the state of a set of nodes at a given horocycle, together with the inputs, provide the state at two descendent sets of nodes at the next horocycle. Since the computations in generating each of these descendent sets are identical in structure, we need follow only one of these paths in order to examine stability. For example, for our sixth-order example, we need only establish stability of the dynamics from input $(E_{61}(m), E_{62}(m))$ to $(Y_1(m), Y_2(m), Y_3(m), Y_4(m))$. However, we can take this considerably farther. In particular, because of the pyramidal symmetries, we need only consider the stability of the map from $(E_{61}(m), E_{62}(m))$ to $Y_1(m)$ as the structure of the map to $Y_2(m)$, $Y_3(m)$, and $Y_4(m)$ are identical. More generally, starting from any node t_0 on the tree, we need only consider the stability of the dynamics involved in generating $\{Y_{t_0\delta^n}, n \ge 0\}$, since the dynamics for any other path from horocycle to horocycle has identical structure.

Using the notation of Fig. 7, we now see that we must examine the stability of the system illustrated in Fig. 8 for the sixth-order case, where the small solid squares now denote standard z^{-1} operations (i.e., $z^{-1}x(m) = x(m-1)$). Here the $S(k_n)$ matrices are exactly as defined in Theorem 3.2. We can now apply standard time domain methods to this system.[4]

Note first that under conditions (4.15), (4.16), the $O(k_n)$ matrices defined in Theorem 3.3 and the covariance matrices Σ_i are invertible so that we can equivalently study the stability of the normalized form of the modeling filter. Note also that checking that a system function $H(z)$ has all its poles strictly inside the unit circle is equivalent to checking the same condition for the system function $H(z^{-1})$. Thus to test for stability we can modify the system of Fig. 8 by adding a unit delay in every left-to-right-going path, and by replacing the $S(k_i)$ blocks by the scattering matrices $\Sigma(k_i)$ of Theorem 3.4. For example, in the sixth-order case we can equivalently check the stability of the system in Fig. 9. Recall that for an Nth-order filter we proved in Theorem 3.4 that

$$\Sigma(k_i)'\Sigma(k_i) = I, \qquad i = 1, \cdots, N \tag{4.17}$$

for any set of coefficients $k_1, \cdots, k_N$ that are reflection coefficients of some isotropic process. But the entries of the matrices $\Sigma(k_i)'\Sigma(k_i) - I$, $i = 1, \cdots, N$ are rational functions of the k_n's that have no poles inside the domain specified by the conditions (4.15), (4.16). Hence we may use the Lemma C.2 of Appendix C to extend the property (4.17) to the whole domain specified by the conditions (4.15), (4.16).

Using (4.17) and the notation of Fig. 9 we have that

$$\|\xi_i(m)\|^2 + \|\eta_i(m)\|^2 = \|\xi_{i-1}(m-1)\|^2 + \|\eta_{i-1}(m-1)\|^2 \tag{4.18}$$

where we have the boundary conditions

$$\xi_{N+1}(m) = u(m) \tag{4.19}$$

$$\eta_0(m) = \xi_1(m-1). \tag{4.20}$$

To study stability we set $u(m) \equiv 0$ and define the follow-

[4]We would like to acknowledge B. C. Levy for suggesting this line of proof.

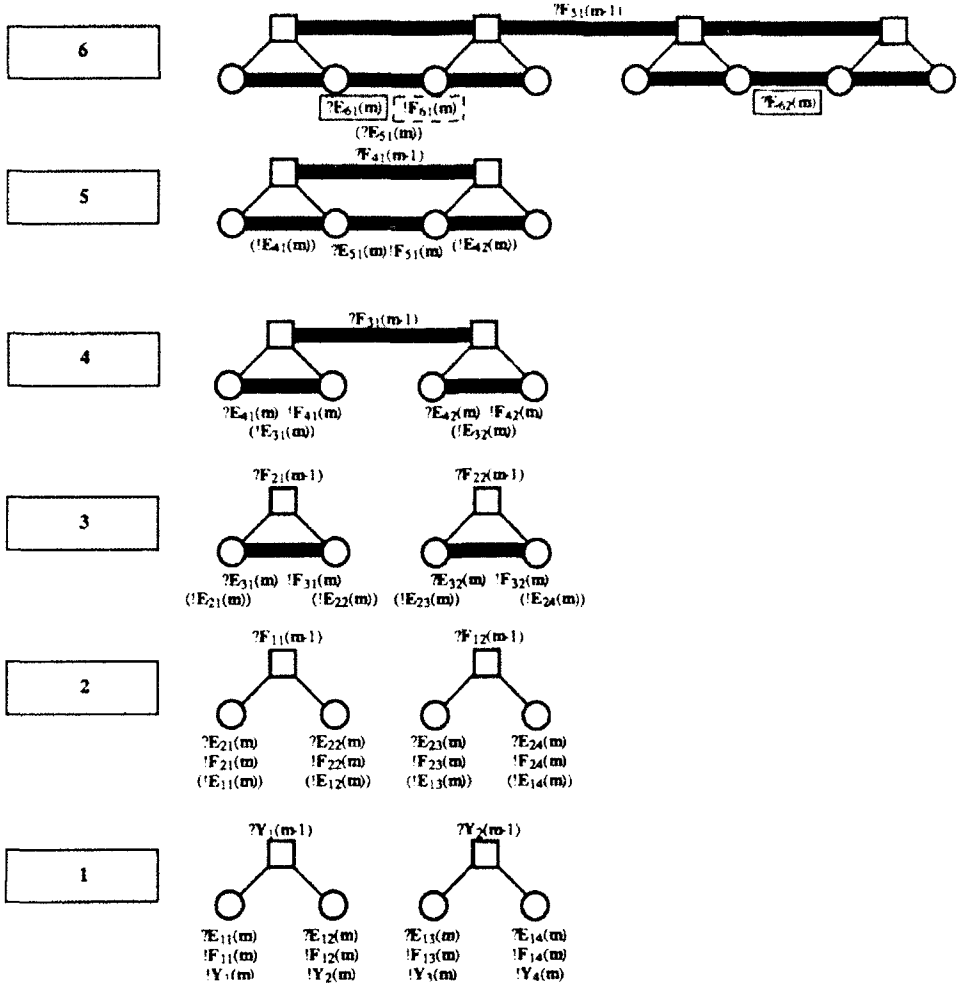

Fig. 7. Illustrating one of the two parallel, sequential computations for the model of Figs. 3–5.

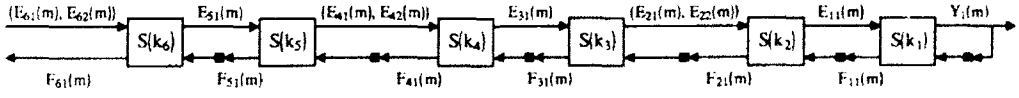

Fig. 8. Illustrating one computational path from horocycle to horocycle. It is this standard time domain system whose stability is equivalent to that of the unnormalized modeling filter.

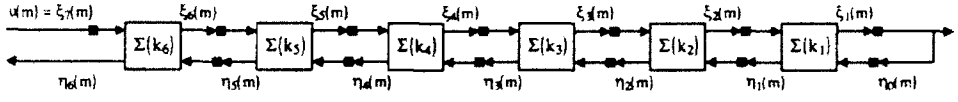

Fig. 9. Equivalent system whose stability is investigated in the proof of Theorem 4.2.

ing positive-definite function of the state of the system

$$V(m) = \sum_{i=1}^{N} \|\xi_i(m)\|^2 + \|\eta_{i-1}(m)\|^2. \qquad (4.21)$$

Then from (4.18)–(4.21) we obtain

$$V(m) - V(m-1) = -\|\eta_N(m)\|^2. \qquad (4.22)$$

It can be readily checked that the system is observable from $\eta_N(m)$, as long as (4.15) and (4.16) are satisfied, so that $V(m)$ is a Lyapunov function proving asymptotic stability.

D. Every Finite Family of Reflection Coefficients Defines an Isotropic AR Process

Our analysis to this point has shown how to construct a sequence of reflection coefficients $\{k_n\}$ from an isotropic covariance sequence $\{r_n\}$. Furthermore, we have seen that the $\{k_n\}$'s have particular bounds and that if $\{r_n\}$ comes from an AR (p) process, only the first p of the reflection coefficients are nonzero. The following result states that the converse holds, i.e., that any finite k_n sequence satisfying the required constraints corresponds to a unique AR covariance sequence. This result substan-

tiates our previous statement that the reflection coefficients provide a good parameterization of AR processes.

Theorem 4.3: Given a finite sequence of reflection coefficients k_n, $1 \le n \le p$ such that

$$\begin{cases} -\frac{1}{2} < k_n < 1 & \text{for } n \text{ even} \\ -1 < k_n < 1 & \text{for } n \text{ odd} \end{cases} \tag{4.23}$$

there exists a unique isotropic covariance sequence which has as its reflection coefficient sequence the given k_n followed by all zeros.

Proof: Consider the modeling filter of order p specified by the given set of reflection coefficients. What we must show is that the output of this filter y_t is well defined (i.e., has finite covariance) and isotropic when the input is a standard white noise process. That it is well-defined follows from the stability result in Theorem 4.2. Thus we need only show that y_t is isotropic. More specifically, let (s, t) and (s', t') be any two pairs of points such that $d(s, t) = d(s', t')$. The theorem will be proved if we can show that the function

$$\Phi: K = (k_n)_{1 \le n \le p} \to E(y_t y_s) - E(y_{t'} y_{s'}) \tag{4.24}$$

is identically zero for all k_n's satisfying the condition (4.23). But the formulas for the modeling filter (Theorem 3.2) show that Φ is a rational function of K which is analytic inside the domain specified by the conditions (4.23). Also Φ is identically zero for all sequences K arising from valid isotropic covariances via the Schur recursions (2.26)–(2.31). Then the theorem is an immediate consequence of the Lemma C.2 of Appendix C.

V. Conclusion

In [1] and this paper we have described a new framework for modeling and analyzing signals at multiple scales. Motivated by the structure of the computations involved in the theory of multiscale signal representations and wavelet transforms, we have examined the class of isotropic processes on a homogenous tree of order 2. Thanks to the geometry of this tree, an isotropic process possesses many symmetries and constraints. These make the class of isotropic autoregressive processes somewhat difficult to describe if we look only at the usual AR coefficient representation. However, as we have developed, the generalization of lattice structures provides a much better parametrization of AR processes in terms of a sequence of reflection coefficients.

In developing this theory we have seen that it is necessary to consider forward and backward prediction errors of dimension that grows geometrically with filter order. Nevertheless, thanks to isotropy, only one reflection coefficient is required for each stage of the whitening and modeling filters for an isotropic process. Indeed as shown in [1], isotropy allowed us to develop a generalization of the Levinson and Schur scalar recursions for the local averages or barycenters of the prediction errors, which also yield the reflection coefficients. In this paper we have justified our claim that the reflection coefficients are a good parametrization for AR processes and isotropic processes in general. In particular we have developed whitening and modeling filters for AR processes that can be completely specified in terms of these coefficients. In addition we have shown that there is a one-to-one correspondence between finite reflection coefficient sequences and AR processes, have characterized the stability of lattice filters in terms of the reflection coefficients and have shown how the regularity of an isotropic process can be characterized in terms of its reflection coefficient sequence.

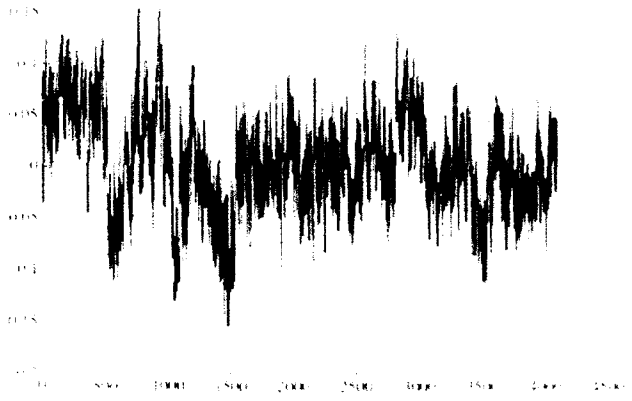

Fig. 10. A sample of an AR (3) process at a given horocycle.

This work represents one part of a larger effort to develop a theoretical foundation for multiscale statistical signal processing. In particular in [10] we investigate a weaker notion of multiscale stationarity which leads to a state space and system theory for multiscale modeling and a corresponding methodology for scale-recursive optimal estimation which accomodates very naturally the fusion of data from sensors with different resolutions [3]–[5]. The multiscale AR models developed here as well as the state-space models of [3]–[5] are particularly useful for modeling and analyzing signals displaying fractal-like or self-similar characteristics. For example, when restricted to a given level of resolution, a sample of an isotropic process can be drawn like an ordinary signal. We show in Fig. 10 a sample of an AR (3) process with $k_1 = k_2 = k_3 = 0.99$. Figs. 11 and 12 show approximations of this signal at successively coarser scales using the multiresolution analysis via wavelets of Mallat-Daubechies, as presented in [7]. These approximations display the self-similar statistical characteristics we expect of this class of models (see also the thorough development for so-called $1/f$-processes in [11], [12]).

There are several promising directions for further research building on our formalism. In particular, an essential topic for investigation is the development of methods for constructing isotropic AR models directly from data as available in practice. This requires identifying multiscale structure and estimating isotropic covariance sequences from data restricted to a single scale, i.e., a single horocycle [6]. In addition, we expect that these models

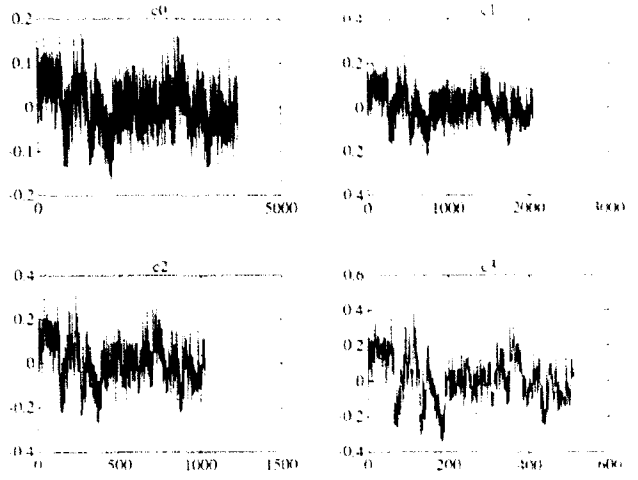

Fig. 11. The Mallat-Daubechies multiresolution approximation of the signal of Fig. 10.

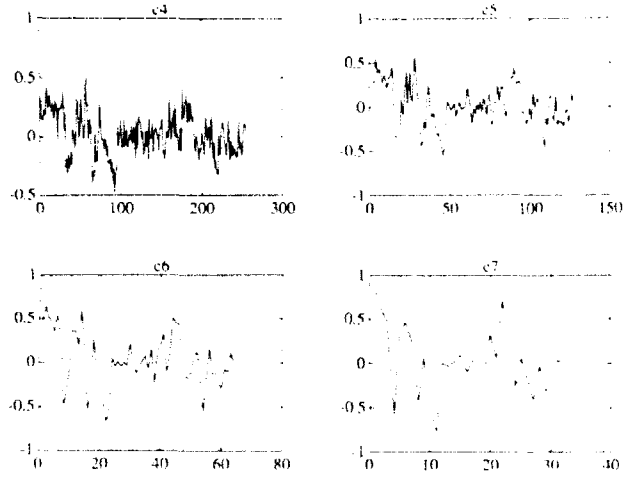

Fig. 12. The Mallat-Daubechies multiresolution approximation of the signal of Fig. 10, continued.

should be of value for segmentation of signals, and, in two dimensions, for the identification of textures, based on differences in multiscale characteristics. The scale-recursive structure of the AR whitening filter should facilitate the calculations of likelihood ratios much as in ordinary time series analysis. Work in these areas, as well as on several applications of our theory, is proceeding and will be reported in the future.

Appendix A
Umbilical Lemma

We shall use the following notation:

$$wY_t = Y_{tw}$$

where w is a word. Note that we have

$$vY_{tw} = Y_{twv} = wvY_t.$$

Furthermore, in the sequel, $[q]$ denotes the greatest integer smaller than q, and we shall write, for short, $\delta^{[q]}$ instead of $\delta^{([q])}$. Using these notations, we have the following result.

Lemma A.1: For each n, the following formulas hold:

$$\gamma_t^{-1}F_{t,n} = \delta^{[(n+1)/2]}\gamma_t^{-1}F_{t,n} \quad \text{up to a permutation.}$$

Proof: Recall that, for $w \le 0$, $|w| = n$

$$F_{t,n}(w) = Y_{tw} - E(Y_{tw} \mid \mathcal{Y}_{t,n-1})$$

whence

$$\gamma_t^{-1}F_{t,n}(w) = Y_{t\gamma_t^{-1}w} - E(Y_{t\gamma_t^{-1}w} \mid \mathcal{Y}_{t,n-1})$$

and

$$\gamma_t^{-1}F_{t\delta^{[(n+1)/2]},n}(w) = Y_{t\delta^{[(n+1)/2]}\gamma_t^{-1}w} - E(Y_{t\delta^{[(n+1)/2]}\gamma_t^{-1}w} \mid \mathcal{Y}_{t\delta^{[(n+1)/2]},n-1})$$

so that, to prove the lemma, it is enough to show the following formulas:

$$|w| = n,\ w' = \delta^{[(n+1)/2]}w \Rightarrow |w'| = n \tag{A.1}$$

$$|w| \le n-1,\ w' = \delta^{[(n+1)/2]}w \Rightarrow |w'| \le n-1. \tag{A.2}$$

Proof of (A.1): Set $w = \gamma_t^{-1}\delta^{(k)}$, $l + 2k = n$. Then

$$w' = \gamma_t^{-1}\delta^{[(n+1)/2]-l}\delta^{(k)} \tag{A.3}$$

where $x_+ = \max(x, 0)$. To prove (A.1) it suffices to verify that

$$k \ge \left[\frac{n-1}{2}\right]_+ - l$$

holds in (A.3), which amounts to verify that $n \ge [(n-1)/2] + k$, and this is a consequence of the inequalities $k \le [n/2]$ and $n \ge [n/2] + [(n-1)/2]$.

Proof of (A.2): Again set $w = \gamma_t^{-1}\delta^{(k)}$, $l + 2k \le n - 1$. Then $w' = \delta^{[(n-1)/2]}w$. Now if $k \ge [(n-1)/2] - l$ holds, then (A.2) follows. Otherwise $l + 2([(n-1)/2] - l) = 2[(n-1)/2] - l \le n - 1$ also proves (A.2).

Appendix B
Proof of (4.4)

Take any $j > n/2$. Suppose that we can find an isometry $\Psi: \mathfrak{I} \to \mathfrak{I}$ so that

1) $\Psi(t) = t$,
2) Ψ maps the set $\{t\gamma_t^{-1}w \mid w \le 0, |w| \le n-1\}$ onto itself,
3) Ψ maps the points $\{t\delta^{(j)}\} \cup \{t\delta^{(j)}\gamma_t^{-1}w \mid w \le 0, |w| \le n-1\}$ onto a set of points each of which is $\le t\gamma_t^{-1}$.

Let $Y_t^\Psi = Y_{\Psi(t)}$ and define E^Ψ similarly. Then, thanks to isotropy, Y^Ψ has the same statistics as Y. Thus from (4.2)

$$E_{t,n}^\Psi \perp \mathcal{Y}_{t\gamma_t^{-1},\infty}^\Psi \tag{B.1}$$

However, thanks to properties (1) and (2) of Ψ

$$E_{t,n} = E_{t,n}^\Psi$$

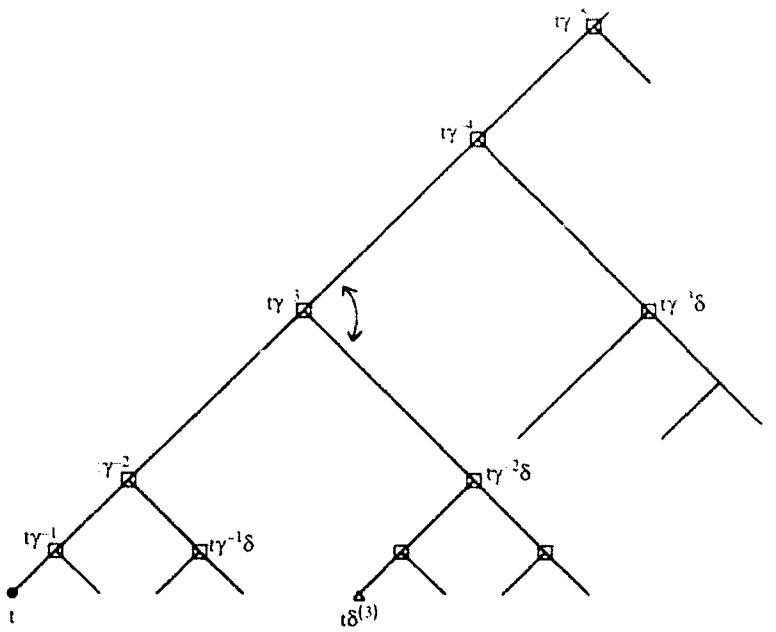

Fig. 13. Illustrating the isometry used in Appendix A for the case $n = 5$ and $j = 3$. Here the pivot point is $t\gamma^{-3}$, so that the part of the tree toward t from $t\gamma^{-3}$ is left unchanged. The "rotation" exchanges the points $t\gamma^{-4}$ and $t\gamma^{-3}\delta$ and maps their successors accordingly. The set of nodes indicated with $\cdot$, which is in this case both $\{t\gamma^{-1}w | w \le 0, |w| \le n-1\}$ and $\{t\delta'\gamma^{-1}w | w \le 0, |w| \le n-1\}$ is left invariant by this isometry. Also the point $t\delta^{(3)}$ is mapped onto one of the immediate successors of $t\gamma^{-3}\delta$, both of which are $\le t\gamma^{-4}$.

while thanks to property (3) and (2.9)

$$E_{t\delta^{(n)},n} \in \mathcal{K}\{Y^{\Psi}_{t\gamma^{-1}w} | w \le 0\}. \tag{B.2}$$

Equations (B.1) and (B.2) then imply (4.4).

The required isometry is of the pivot type used in the proofs in [1, appendix C]. As illustrated in Fig. 13, the pivot for this isometry is the point $t\gamma^{-(1+n/2)}$ and the direction of "rotation" is as indicated in the figure. It is straightforward to check that this isometry has the required properties.

Appendix C
Some Useful Lemmas

The first lemma is an immediate consequence of the Schur recursions (22.6)–(2.31):

Lemma C.1: Consider the transformation Ψ which maps an isotropic covariance sequence $\{r_n\}$ to the corresponding reflection coefficient sequence. The Jacobian of this transformation satisfies the following:

$$\frac{\partial k_n}{\partial r_m} = 0 \qquad \text{for } n < m \tag{C.1}$$

$$\frac{\partial k_{2n}}{\partial r_{2n}} = \frac{1}{2^{n-1}P_{2n-1}(0)} \neq 0 \tag{C.2}$$

$$\frac{\partial k_{2n+1}}{\partial r_{2n+1}} = \frac{1}{2^{n-1}(P_{2n}(0) + \delta^{(n)}|P_{2n}|(0))} \neq 0 \tag{C.3}$$

where the P_n are the Schur series defined in (2.26).

Next we write $K = (k_n)_{1 \le n \le p}$ to denote a vector in R^p, and we let $\mathcal{S}$ denote the set of such vectors so that

$$-1 < k_{2n+1} < +1$$

$$-\tfrac{1}{2} < k_{2n} < +1.$$

Lemma C.2: Consider a function Φ from R^p into R satisfying the following properties:

1) $\Phi(K) = 0$ if K is the reflection coefficient sequence of an isotropic process.

2) Φ is analytic inside $\mathcal{S}$.

Then, $\Phi \equiv 0$ in $\mathcal{S}$.

Proof: Since Φ is analytic in $\mathcal{S}$, it is sufficient to show that Φ is zero on a set with nonempty interior in $\mathcal{S}$. Since we know that $\Phi(K) = 0$ if K is in the image of the map Ψ introduced in Lemma C.1, it is sufficient for us to show that the image of Ψ has a nonempty interior.

Thanks to the form of the Schur recursion formulae (2.26)–(2.31), we know that Ψ is also a rational function and, thanks to Lemma C.1, its Jacobian is triangular and always invertible. Thus it is sufficient to show that the set of finite sequences $\{r_n | 0 \le n \le N\}$ that can be extended to a covariance function of an isotropic process has a nonempty interior. However, this property is characterized by a finite family of conditions of the form

$$\mathcal{R}(r_0, \cdots, r_p) \ge 0 \tag{C.4}$$

where $\mathcal{R}(r_0, \cdots, r_p)$ denotes a matrix whose elements are chosen from the $r_0, \cdots, r_p$. The set of $(p+1)$-tuples satisfying these conditions with strict inequality is nonempty (for instance, $r_n = \delta_{n0}$ is the covariance sequence of white noise) and as a consequence the set of r_0,

$\cdots$, r_p satisfying (C.4) has a nonempty interior. This finishes the proof of the lemma.

Acknowledgment

The authors would like to express their thanks to O. Zeitouni, R. Nikoukhah, and especially to B. C. Lévy for extremely fruitful discussions and stimulating suggestions. They are also grateful to the reviewers for their many constructive comments for this paper and its companion.

The research of A. S. Willsky was performed while he was a visitor at INRIA.

References

[1] M. Basseville, A. Benveniste, and A. S. Willsky, "Multiscale autoregressive processes, Part I: Schur-Levinson parametrizations," this issue, pp. 1915-1934.

[2] A. Benveniste, "Methodes d'orthogonalisation en treillis pour le problème de la réalisation stochastique," in *Outils et Modèles Mathématiques pour l'Automatique, l'Analyse de Systèmes et le Traitement du Signal*, vol. 2. Editions du CNRS, 1982.

[3] K. C. Chou, A. S. Willsky, A. Benveniste, and M. Basseville, "Recursive and iterative estimation algorithms for multiresolution stochastic processes," in *Proc. IEEE Conf. Decision Contr.* (Tampa, FL), Dec. 1989.

[4] K. Chou, "A stochastic modeling approach to multiscale signal processing," Ph.D. disseration, Dep. Elect. Eng. Comput. Sci., M.I.T., 1991.

[5] K. C. Chou and A. S. Willsky, "Multiscale Riccati equation and a two-sweep algorithm for the optimal fusion of multiresolution data," in *Proc. 29th IEEE Conf. Decision Contr.* (Honolulu, HI), Dec. 1990.

[6] B. Claus, "Model fitting for multiscale isotropic processes," Ph.D. dissertation, IRISA, Rennes, France, in preparation.

[7] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 91, pp. 909-996, 1988.

[8] C. C. Heyde and D. Hall, *Martingale Limit Theory and Application*. New York: Academic, 1980.

[9] J. Neveu, *Martingales à temps discret*. Paris, France: Masson, 1972.

[10] A. Benveniste, R. Nikoukhah, and A. S. Willsky, "Multiscale system theory," in *Proc. 29th IEEE Conf. Decision Contr.* (Honolulu, HI), Dec. 1990.

[11] G. W. Wornell, "A Karhunen-Loeve like expansion for 1/*f* processes via wavelets," *IEEE Trans. Inform. Theory*, vol. 36, no. 9, pp. 859-861, July 1990.

[12] G. W. Wornell and A. V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets," *IEEE Trans. Signal Processing*, vol. 40, no. 3, pp. 611-623, Mar. 1992.

Michèle Basseville was born on May 2, 1952 in Paris, France. In 1976, she graduated from Ecole Normale Supérieure de Fontenay-aux Roses, where she studied mathematics, and she received the Docteur-es-Sciences degree (thèse d'Etat) in 1982.

Since 1976, she has been with the Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), where she held a CNRS position, "Chargée de Recherche," from 1976 until 1990, and where she is now Directeur de Recherche at CNRS. Her main domain of interest has been the detection of abrupt changes in signals and dynamical systems, and its application to both recognition-oriented signal processing and industrial process monitoring for conditional maintenance. In 1982, she obtained the Docteur-es Sciences degree (thèse d'Etat) for her contribution to several on-line signal segmentation algorithms, which have been proved useful since then, especially for geophysical signal processing and continuous speech recognition. Since 1980, she and A. Benveniste have been involved in a joint project on vibration monitoring for both large mechanical structures and rotating machines. Since 1988, they have been investigating other possible applications for monitoring, among them gas turbines. She has edited together with A. Benveniste the collective monograph *Detection of Abrupt Changes in Signals and Dynamical Systems* (Lecture Notes in Control and Information Sciences no. 77). She is currently writing a book, *Detection of Abrupt Changes: Theory and Applications*, with I. V. Nikiforov (IPOI, Moscow), which is to appear in the Prentice Hall Information and System Sciences Series. Since 1987, she has been involved in the joint project INRIA-MIT on a statistical theory for multiresolution analysis of signals and images, with A. Benveniste and A. Willsky. Her main interest in this project is the development of statistical algorithms for the detection and recognition of events in multiscale signal processing.

Albert Benveniste (M'81–SM'89–F'92) was born May 8, 1949 in Paris, France. He graduated from Ecole des Mines de Paris in 1971 and received the These d'Etat degree in mathematics (probability theory) in 1975.

From 1976 to 1979 he was Associate Professor of Mathematics at Université de Rennes I. Since 1979 he has been Directeur de Recherche at INRIA. For his thesis he worked on probability theory, stochastic processes, and ergodic theory. In parallel, he pursued work on automatic control and signal processing in the area of adaptive systems for time-varying systems with applications to data communications. In 1980 he began joint work with M. Basseville on the subject of change detection in signals and dynamical systems with application to vibration mechanics and fault detection in process control. Since 1987, he has been involved in developing, jointly with A. S. Willsky, a statistical theory of multiresolution signal and image processing, based on dynamical systems and Gaussian random fields on homogeneous trees. Since 1981, he has been interested in computer science, in the area of real-time languages and systems. Cooperating with P. Le Guernic, he participated in the definition of and theoretic studies on the Signal language.

Dr. Benveniste was corecipient in 1980 of the IEEE Transactions on Automatic Control Best Transaction Paper Award for his paper on blind deconvolution in data communications. In 1990 he received the CNRS Silver Medal. From 1986 to 1990 he was Vice Chairman of the IFAC Committee on Theory and is Chairman of this committee for 1991–1993. From 1987 to 1990 he was Associate Editor for IEEE Transactions on Automatic Control. He is currently Associate Editor for the *International Journal of Adaptive Control and Signal Processing*, the *International Journal of Discrete Event Dynamical Systems*, and Associate Editor at Large for the IEEE Transactions on Automatic Control. He has coauthored with M. Metivier and P. Priouret the book *Adaptive Algorithms and Stochastic Approximations*, and has been an Editor, with M. Basseville, of the collective monograph *Detection of Abrupt Changes in Signals and Systems*.

Alan S. Willsky (S'70–M'73–SM'82–F'86) received the S.B. and Ph.D. degrees from the Massachusetts Institute of Technology in 1969 and 1973, respectively.

From 1969 through 1973 he held a Hannie and John Hertz Foundation Fellowship. He joined the M.I.T. faculty in 1973 and his present position is Professor of Electrical Engineering. From 1974 to 1981 he served as Assistant Director of the M.I.T. Laboratory for Information and Decision Systems. He is also a founder and Member of the Board of Directors of Alphatech, Inc. He has held visiting positions at Imperial College, London, l'Universite de Paris Sud, and the Institut de Recherche en Informatique et Systemes Aleatoires in Rennes, France. He is Editor of the M.I.T. Press series on signal processing, optimization, and control. His present research interests are in problems involving multidimensional and multiresolution estimation and imaging, discrete event systems, and the asymptotic analysis of control and estimation systems. He is

the author of the research monograph *Digital Signal Processing and Control and Estimation Theory* and is coauthor of the undergraduate text *Signals and Systems*.

Dr. Willsky was Program Chairman for the 17th IEEE Conference on Decision and Control, has been an Associate Editor of several journals, including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, has served as a Member of the Board of Governors and Vice President for Technical Affairs of the IEEE Control Systems Society, and was Program Chairman for the 1981 Bilateral Seminar on Control Systems held in the People's Republic of China. He is also the Special Guest Editor of the 1992 special issue of the IEEE TRANSACTIONS ON INFORMATION THEORY on wavelet transforms and multiresolution signal analysis. In 1988 he was made a Distinguished Member of the IEEE Control Systems Society. In addition, he has given several plenary lectures at major scientific meetings, including the 20th IEEE Conference on Decision and Control and the 1991 IEEE International Conference on Systems Engineering. In 1975 he received the Donald P. Eckman Award from the American Automatic Control Council. He was awarded the 1979 Alfred Noble Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE for a paper excerpted from his monograph.